

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:46:10 ; Search time 12 seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-9

Perfect score: 31

Sequence: 1 WXXWXP 6

Scoring table: BLOSUM62DX

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	100.0	21	2	PH1731	Ig heavy chain V r
2	31	100.0	49	2	I48681	185 kDa glycoposp
3	31	100.0	59	2	E82844	hypothetical prote
4	31	100.0	61	2	H91196	hypothetical prote
5	31	100.0	65	2	S23164	light-harvesting p
6	31	100.0	71	2	F84168	hypothetical prote
7	31	100.0	73	2	AH2788	hypothetical prote
8	31	100.0	76	2	T14724	hypothetical prote
9	31	100.0	78	2	S08449	hypothetical prote
10	31	100.0	79	2	A64750	hypothetical prote
11	31	100.0	83	2	A97564	conserved hypothet
12	31	100.0	86	2	T42185	prophage p10 prote
13	31	100.0	87	2	H86753	Ig kappa chain V r
14	31	100.0	92	2	S37509	hypothetical 10.7K
15	31	100.0	92	2	T21627	hypothetical prote
16	31	100.0	93	2	G87098	hypothetical prote
17	31	100.0	97	2	E84147	hypothetical prote
18	31	100.0	99	2	C84022	hypothetical prote
19	31	100.0	101	2	S72519	frbabilis lipoprote
20	31	100.0	103	2	F82266	conserved hypothet
21	31	100.0	105	2	JEC384	NADH2 dehydrogenas
22	31	100.0	106	2	T64171	hypothetical prote
23	31	100.0	106	2	C82715	conserved hypothet
24	31	100.0	106	2	G72022	transport permease
25	31	100.0	108	2	AG1164	hypothetical prote
26	31	100.0	108	2	S28241	NADH2 dehydrogenas
27	31	100.0	108	2	AC2505	hypothetical prote
28	31	100.0	109	2	F83057	hypothetical prote
29	31	100.0	109	2	AT2029	hypothetical prote

30	31	100.0	110	2	AD0127	conserved hypothet
31	31	100.0	110	2	AG1523	hypothetical prote
32	31	100.0	116	2	P28966	Ig heavy chain pre
33	31	100.0	116	2	G82398	conserved hypothet
34	31	100.0	117	2	A95381	hypothetical prote
35	31	100.0	118	2	S24527	Ig heavy chain V r
36	31	100.0	119	2	S24522	Ig heavy chain V r
37	31	100.0	119	2	S24422	Ig heavy chain V r
38	31	100.0	119	2	S24513	Ig heavy chain V r
39	31	100.0	119	2	S24493	Ig heavy chain V r
40	31	100.0	119	2	S24491	Ig heavy chain V r
41	31	100.0	119	2	S24525	Ig heavy chain V r
42	31	100.0	119	2	S24520	Ig heavy chain V r
43	31	100.0	119	2	S24519	Ig heavy chain V r
44	31	100.0	119	2	S24511	Ig heavy chain V r
45	31	100.0	119	2	S24424	Ig heavy chain V r

ALIGNMENTS

RESULT 1
PH1731
Ig heavy chain V region (clone GCC-11) - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 24-Feb-1994 #sequence_revision 24 Feb 1994 #text_change 17-Mar-1999
C/Accession: PH1731
R/McHeyzer Williams, M.C., McLean, M.J., Lalor, P.A., Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A/Title: Antigen-driven B cell differentiation in vivo.
A/Reference number: PH1675; MIMD:93301607; PMID:8115185
A/Accession: PH1731
A/Molecule type: mRNA
A/Residues: 1-21 <MCH>
A/Experimental source: B cell
A/Note: the authors translated the codon ACA for residue 13 as Ala
C/Superfamily: immunoglobulin V region, immunoglobulin homology
C/Keywords: heterotetramer, immunoglobulin

Query Match 100.0%; Score 31; DB 2; Length 21;
Best Local Similarity 50.0%; Pred No 1 30+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
DB 16 WTSWSP 21

RESULT 2
I48681
185 kDa glycoposphoprotein - mouse (fragment)

C/Species: Mus musculus (house mouse)
C/Date: 02-Jul-1996 #sequence_revision 02-Jul 1996 #text_change 02-Jun 2002
C/Accession: I48681
P/White, M.R., Hung, M.C.
Oncogene 7, 677-683, 1992
A/Title: Cloning and characterization of the mouse neu promoter.
A/Reference number: I48681, MIMD:92284966; PMID:1348855
A/Accession: I48681
A/Status: preliminary; translated from GR/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-49 <RES>
A/Cross-references: EMBL: X66236, MIMD:9335196, FIDN: CAA46266.1, FID: 3195187
C/Genetics:
A/Gene: neu
C/Keywords: ATP; phosphoprotein

Query Match 100.0%; Score 31; DB 2; Length 49;
Best Local Similarity 50.0%; Pred No 2 70+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
DB 16 WTSWSP 21

A/Accession: AH2788
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-77 <YRP>
 A/Cross-references: GB:AF006664; FID:AA147226.1; FID:g17749165; USFCF:JN00186
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A/Gene: Atul727
 A/Map position: circular chromosome

Query Match 100.0%; Score 31; DB 2; Length 73;
 Best Local Similarity 50.0%; Pred. No. 4e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
 |||||
 Db 13 WNLWIF 18

RESULT 8

hypothetical protein Y1001 - Yersinia pestis plasmid pMT1
 C/Species: Yersinia pestis
 C/Date: 20-Sep-1999 #sequence_revision 20 Sep 1999 #text_change 24-Sep-1999
 C/Accession: T14724; T15006
 R/Hu, P. Elliott, C. McCarthy, P. Skowronski, E. Garner, C. Kobayashi, A. Carrano,
 submitted to the EMBL Data Library, March 1999
 A/Description: Structural organization of virulence determinants in three Yersinia pesti
 A/Reference number: 218168
 A/Accession: T14724
 A/Status: preliminary; translated from GP/EMBL/PPPT
 A/Molecule type: DNA
 A/Residues: 1-76 <HUP>
 A/Cross-references: EMBL:AF053947, MIP:G0992296, FID:G0996363, FID:AA013241;
 R/Lindler, L.E.; Plano, G.V.; Rutland, V.; Mayhew, G.F.; Blattner, F.P.
 Infect. Immun. 66, 5731-5742, 1998
 A/Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIMS plasmid
 A/Reference number: 218268; MIP:00043809; PMID:9826348
 A/Accession: T15006
 A/Status: preliminary; translated from GP/EMBL/PPPT
 A/Molecule type: DNA
 A/Residues: 1-76 <HUP>
 A/Cross-references: EMBL:AF074611, MIP:G3993003, FID:G3993009, FID:AA082749;
 C/Genetics:
 A/Gene: Y1091
 A/Genome: plasmid pMT1

Query Match 100.0%; Score 31; DB 2; Length 76;
 Best Local Similarity 50.0%; Pred. No. 4e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
 |||||
 Db 50 WRTWNF 55

RESULT 9

S08449
 hypothetical protein 3 Spiroplasma virus 1
 C/Species: Spiroplasma virus 1, Spv1
 C/Date: 10-Jun-1991 #sequence_revision 10-Jun-1991 #text_change 07-Dec-1999
 C/Accession: S08449
 R/Renandrin, J.; Aulio, P.; Vignault, J.C.; Bove, J.M.
 Nucleic Acids Res. 18, 1993, 1999
 A/Title: Complete nucleotide sequence of the genome of Spiroplasma citri virus Spv1-R9A2
 A/Reference number: S08447; MIP:0026799; PMID:2329423
 A/Accession: S08449
 A/Status: translation not shown
 A/Molecule type: DNA
 A/Residues: 1-78 <REN>
 A/Cross-references: EMBL:X61344; MIP:g151003; FID:CAAT5777.1; FID:G064399
 C/Genetics:
 A/Genetic code: SGC3

Query Match 100.0%; Score 31; DB 2; Length 78;
 Best Local Similarity 50.0%; Pred. No. 4.1e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
 |||||
 Db 42 WLIWLF 47

RESULT 10

hypothetical protein b0249 - Escherichia coli (strain K-12)
 C/Species: Escherichia coli
 C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C/Accession: A64750
 R/Blattner, F.P.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Rutland, V.; Riley, M.; Gill
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MIP:07426417; PMID:9278603
 A/Accession: A64750
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-79 <BLAT>
 A/Cross-references: GB:AF06133, G0100005, FID:G067699, FID:AA000000.1; FID:G0671137
 A/Experimental source: strain K-12, substrain MG1655

Query Match 100.0%; Score 31; DB 2; Length 79;
 Best Local Similarity 50.0%; Pred. No. 4.1e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
 |||||
 Db 50 WPAWNP 55

RESULT 11

hypothetical protein AGP_C_3173 (imported) Agrobacterium tumefaciens (strain C58, Cere
 C/Species: Agrobacterium tumefaciens
 C/Date: 30-Sep-2001 #sequence_revision 30 Sep 2001 #text_change 11-Jan-2002
 C/Accession: A97568
 R/Goodner, R.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Golden,
 A.; Liu, F.; Wellam, C.; Allinger, M.; Deighy, D.; Scott, C.; Tappin, C.; Matvelis, B.;
 Science 294, 2223-2228, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A/Reference number: A97559; PMID:11743194
 A/Accession: A97568
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-81 <KUR>
 A/Cross-references: GB:AE007869; FID:AA01408.1; FID:g1516475; G099999919
 C/Genetics:
 A/Gene: AGR_C_3173
 A/Map position: circular chromosome

Query Match 100.0%; Score 31; DB 2; Length 83;
 Best Local Similarity 50.0%; Pred. No. 4.3e+02;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
 |||||
 Db 23 WNLWIF 28

RESULT 12

T42185
 conserved hypothetical protein L7085 - Escherichia coli plasmid p0157
 C/Species: Escherichia coli
 C/Date: 02-Dec-1999 #sequence_revision 02-Dec-1999 #text_change 03-Dec-1999
 C/Accession: T42185
 R/Rutland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.D.; Blattner, F.P.

GenInfo version 5.1.3
Copyright (c) 1993 2003 CompuGen Ltd

OM protein - protein search, using SW model

Run on: February 20, 2003, 10:41:55, Search time 6.6667 seconds
(without alignments)
37.329 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 31
Sequence: 1 WXXWXP 6

Scoring table: BLOSUM62DX
Gapop 10.0, Gapext 0.5

Searches: 119992 seqs, 41476328 residues

Total number of hits satisfying chosen parameters 119992

Minimum DB seq length: 0
Maximum DB seq length: 2600000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	31	100.0	65	1HA2_ECTHL	P80103 Ectothiorhodospira halochloris
2	31	100.0	78	VG9_SPVIR	P15600 Spiroplasma
3	31	100.0	79	YKFF_ECOLI	P75677 Escherichia
4	31	100.0	84	ACHA_CROPS	P54248 Crocidura
5	31	100.0	84	ACHA_FELCA	P54250 Felis silve
6	31	100.0	90	YPR1_ECOLI	P58095 Escherichia
7	31	100.0	101	YGS0_MYCPN	P75147 Mycoplasma
8	31	100.0	105	NIGM_HUMAN	P95178 Homo sapien
9	31	100.0	106	YGCC_HAFIN	P97152 Haemophilus
10	31	100.0	109	NISM_PAVIN	P92754 Bos taurus
11	31	100.0	116	HV06_CARAU	P19181 Carassius a
12	31	100.0	128	MERT_STAU	P08656 Staphylococ
13	31	100.0	129	YV3H_HUMAN	P04207 Homo sapien
14	31	100.0	129	YIMW_YEAST	P38957 Saccharomyc
15	31	100.0	138	YAA1_MYCPN	P75602 Mycoplasma
16	31	100.0	139	YEL3_MYCPN	P95x45 Mycoplasma
17	31	100.0	144	HV26_MOUSE	P01795 Mus musculu
18	31	100.0	145	P132_AEPPE	P95f02 Acropora
19	31	100.0	146	YGS6_YEAST	P79311 Saccharomyc
20	31	100.0	147	YSE6_METJA	P57095 Methanococ
21	31	100.0	155	H-PE1_ECOLI	P08932 Escherichia
22	31	100.0	157	YEE3_MYCPN	P75320 Mycoplasma
23	31	100.0	164	LSPA_ECOLI	P00904 Escherichia
24	31	100.0	165	DSBR_VIBAL	P56578 Vibrio algi
25	31	100.0	165	LSPA_ENTAS	P13514 Enterobacte
26	31	100.0	165	YPR1_PHOSH	P53229 Rhodobacte
27	31	100.0	166	Y19K_SSV1	P20209 Sulfolobus
28	31	100.0	170	LSPA_PSEPL	P17942 Pseudomonas
29	31	100.0	171	Y428_MYCGE	P47667 Mycoplasma
30	31	100.0	172	YB03_MYCPN	P75566 Mycoplasma
31	31	100.0	176	VPI_BPP2	P26701 Bacterioph
32	31	100.0	177	TN18_HUMAN	P99932 Homo sapien
33	31	100.0	182	VG37_BPMIT	P91198 Bacterioph

34	31	100.0	196	YJPR_YEAST	P46989 Saccharomyc
35	31	100.0	200	YK01_MYCTY	P71669 Mycobacteri
36	31	100.0	206	YVIE_ECOLI	P75707 Escherichia
37	31	100.0	208	YHNN_ECOLI	P37616 Escherichia
38	31	100.0	210	SSH6_YEAST	P01446 Saccharomyc
39	31	100.0	213	IF4E_XENLA	P48537 Xenopus lae
40	31	100.0	215	IF4E_APLIC	P77210 Aplysia cal
41	31	100.0	217	IF4E_HUMAN	P66710 Homo sapien
42	31	100.0	217	IF4E_MOUSE	P29417 Mus musculu
43	31	100.0	217	IF4E_PABIT	P29319 Crystallatus
44	31	100.0	219	FMGF_METJA	P57768 Methanococ
45	31	100.0	220	RP28_BPSPI	P03048 Bacterioph

ALIGNMENTS

RESULT 1
ID LHA2_ECTHL STANDARD: PRT: 65 AA.
AC P80103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB 1994 (Rel. 28, Last annotation update)
DE Light-harvesting protein P800/840/1020, alpha-2 chain (F8S-alpha 2)
DE (Antenna pigment protein, alpha-2 chain).
OS Ectothiorhodospira halochloris.
OC Bacteria; Proteobacteria; gamma subdivision; Ectothiorhodospiraceae;
OC Halorhodospira.
CY NCBI_TaxID 1052;
RN (1)
RP SEQUENCE.
RC STRAIN=DSM 1059;
FX MEDLINE=92249336; PubMed=1577009;
RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
Zuber H.;
RA "The primary structure of the antenna polypeptides of
ET Ectothiorhodospira halochloris and Ectothiorhodospira halophila. Four
ET core-type antenna polypeptides in E. halochloris and E. halophila.";
RL Eur. J Biochem 205:917-925(1992).
CC - FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC - SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON PROMOTER GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC - SUPPLEMENTARY LOCATION: Type II membrane protein. Inner membrane.
CC FIR, 323164; S23164.
CC InterPro: IPR002361; AntennaA.
CC InterPro: IPR000066; Antenna_a/b.
CC Pfam: PF00556; LHC1.1.
CC PROSITE, P00768; ANTENNA_COMP_ALPHA; 1.
CC Antenna complex, light harvesting polypeptide, transmembrane;
KW Magnesium; Bacteriochlorophyll; inner membrane.
KW DOMAIN
FT 1
FT TRANSMEM 14 34
FT DOMAIN 35 65
FT METAL 23 29
FT METAL 29 29
FT SEQUENCE 65 AA; 7688 MW, 55443.96748339 DA (P03044);
Query Match 100.0%; Score 31; P1; Length 65;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 WXXWXP 6
DB 2 WKLWKF 7
RESULT 2
VG9_SPVIR

ID VGG SPITE STANDARD, RT, TS AA.
 AC P15900;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 CE Gene 3 protein.
 GN 3
 OS Spiroplasma virus CPY1 RNA 3.
 OS Viruses, RNA viruses, Spiroplasma, Spiroplasma.
 CC NCBI TaxID=10854;
 RX 11)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90206789; PubMed=2320423;
 SA Renardin M, Audin R, Vignat B, Roux M
 RT "Complete nucleotide sequence of the genome of Spiroplasma CPY1
 RT virus (SEVI-PAC RNA)".
 RL Nucleic Acids Res. 18:1293-1293(1990).
 CC -----
 CC This Swiss spot entry is copyright. It is provided through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL database.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Single by and for commercial
 CC entities requires a licensing agreement. More help is available at: <http://www.ebi.ac.uk/annotation/>
 CC or send an email to license@isb.sib.ch.

Query Match:	100.0%;	Score 31;	DB 1;	Length 78;
Best Local Similarity	50.0%;	Pred. No. 1.6e+02;		
Matches	2;	Conservative	0;	Mismatched 0;

2y	1	XXXXXX	6
		1:1:1	
Db	42	WLTLPL	47

```

1  RESULT 3
2  YKFF_ECOLI
3
4  STANDARD;          PRT;          79 AA
5
6  15-JUL-1998 (Rel. 36, Created)
7  15-JUL-1998 (Rel. 36, Last sequence update)
8  16-OCT-2003 (Rel. 43, Last modification)
9  Hypothetical protein ykff
10 YKFF OR F0249.
11 Escherichia coli
12 Bacteriophage, Prophage, gamma subphage, Enterb-1998.
13 Escherichia
14 NCBI_TaxID=562;
15 (1)
16
17 SEQUENCE FROM N.A.
18 STRAIN=K12 / MS1655;
19 MEDLINE 97416817, PubMed 9273903;
20 Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
21 Riley M., Collado-Vides J., Glasner J.S., Rode C.K., Mayhew D.W.,
22 Gregor T., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
23 Mau B., Shao Y.;
24 "The complete genome sequence of Escherichia coli K-12."
25 Science 277:1493-1474(1997).
26
27 SEQUENCE FROM N.A.
28 STRAIN=K12 / W3110;
29 Takemoto Y., Mori H., Murayama N., Nakajima K., Yano M., Ishii T.,
30 Yamamoto Y., Inohara H., Miyai T., Harada E., Ikeda R., Ishihara G.,
31 Mizuno T., Makino K., Nakata A., Yama T., Carpel G., Mikami K.,
32 "Systematic sequencing of the Escherichia coli genome. Analyses of the
33 4.6 - 6.0 min (189,987 - 291,416bp) region."
34 Submitted (FEB 1996) to the EMBL/GenBank/CCRF databases
35 -1- SIMILARITY: STRONG, to E. coli YP01.
36
37 THIS SWISS PROT entry is copyright, it is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions in any of its content in any way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.ebi.ac.uk/awards>
CC or send an email to license@ebi.ac.uk.

LA EMERGENCY MEDICAL CENTER, 600 E. 9TH AVE.,
OKLAHOMA CITY, OKLAHOMA 73104-0800
DR. JAMES H. HARRIS, JR., M.D., F.A.C.P.
KIMBERLY HARRIS, R.N., C.R.N.P.
COUNCIL ON AGING, 100 N. WILSON ST., SUITE 200,
OKLAHOMA CITY, OKLAHOMA 73104

Empty Match	100.0%	Score 317	DB 17	Length 79;
Full Match	100.0%	Indel 100.00e+02;		
Mismatches	0	Consecutive	0	Indels
			0	Taps

OY	1	WXXWF	6
EB	50	WFAWNE	55

```

RESULT 4
ACHA_CROSS
ID_ACHA_CROSS      STANDARD;      PRT;      84 AA

```

IT 11-0001-1496 (Prel. 34, Created)

17	1997-1998 (Feb. 94)	Last sequence update)
DT	15 JUN 2002 (Feb. 41)	Last annotation update)

DE Acetylcholine receptor protein, alpha chain (

White footed shrew).

Mammalia; Eutheria; Insectivora; Crocodyriformes; Crocodyrura.

$$\mathbf{r}_N = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

RC TISSUE: Muscle;

RA Barclan D., Ovadia M., Kochva E., Fuchs S.;

species resistant to alpha-bungarotoxin."

1. ETHIOPIA: AFTER THE DEATH OF THE KING, THE ETHIOPIAN GOVERNMENT HAS BEEN TRYING TO RECOVER THE COUNTRY FROM THE HANDS OF THE ARABIAI.

[illegible][illegible]

CC MUSCLE) CHAINS.
CC SURCELLULAR LOCATION: Integral membrane protein (By similarity)

-1- SIMILARITY. BELONGS TO THE GROUP CALLED IONIC CHANNEL FAMILY.

CC This SWISS-ELECT entry is copyright © 1997 by the EMBL Laboratory

the European Microinformatics Institute. There are no restrictions on

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

02 500 7 3: 6647 40 1100000100 536.20).

DR EMRL; U17006; AAA89113.1; -.
Interbro. IF001175. Naur channel]

Plam; PF02931; Neur chan LBD; 1.

14. *Receptor, histophoric, cell-surface, glycoprotein, transmembrane*

RE: STEPHEN R. STEINBERG

[illegible][illegible]

Downloaded from <http://ajph.org/> on November 10, 2015

SEQUENCE 84 AA; 6833 MW; 732072PBO967096P CPO664;

Query Match 100.0%; Score 31; DB 1; Length 84;

Best Local Similarity 50.0%; Pred No 1 7a+02;

Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6

DB 63 WKHWVF 68

RESULT 5

ACHA_FEICA

ID_ACHA_FEICA STANDARD; PRT; 84 AA.

AC P54250;

DT 01-OCT-1996 (Rel 34, Created)

DT 01-OCT-1996 (Rel 34, Last sequence update)

DT 15-JUN-2002 (Rel 41, Last annotation update)

DE Acetylcholine receptor protein, alpha chain (fragment)

GN CHRNA1.

OS Felis silvestris catus (Cat).

OC Eukaryota; Metazoa; Chordata; Carnivora; Felidae; Felis.

OC Mammalia; Euthera; Carnivora; Fissipedia; Felidae; Felis.

OX NCBI_Taxid=9685;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Muscle;

RX MEDLINE=95345091; PubMed=7619817;

RA Baraban D, Ovadia M, Kochva E, Fuchs S;

RT "The binding site of the nicotinic acetylcholine receptor in animal

species resistant to alpha-bungarotoxin";

RL Biochemistry 34:9172-9176(1995).

CC -1- FUNCTION: ACETYLCHOLINE, THE ACHP RESPONSES BY AN

EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND

LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA

MEMBRANE.

CC -1- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA,

DELTA, AND GAMMA (IN IMMATURE MUSCLE) OR Epsilon (IN MATURE

MUSCLE) CHAINS.

CC -1- SUPPLEMENTAL LOCATION: Integral membrane protein (By similarity).

CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U17007; AAA89114.1;

DR InterPro; IPR001175; Neur_chan.

DR Pfam; PF02931; Neur_chan_LBD; 1.

DR PROSITE; PS00235; MEMPTP_ION_CHANNEL; 1.

KW Receptor; Postsynaptic membrane; Ion channel; Glycoprotein;

Transmembrane.

FT NON_TER 1 1

FT DISULFID 7 21

FT DISULFID 71 72

FT CARPOHYD 20 20

FT NON_TER 84 84

SO SEQUENCE 84 AA; 9823 MW;

SEAFVAVLQVAFPRVVF CPO64;

Query Match 100.0%; Score 31; DB 1; Length 84;

Best Local Similarity 50.0%; Pred No 1 7a+02;

Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6

DB 63 WKHWVF 68

RESULT 6

YPJ1_ECOLI

ID_YPJ1_ECOLI STANDARD; PRT; 90 AA.

AC P58095;

DT 16-OCT-2001 (Rel 40, Created)

DT 16-OCT-2001 (Rel 40, Last sequence update)

DT 15-JUN-2002 (Rel 41, Last annotation update)

DE Hypothetical protein YPJ1.

GN YPJ1 OR R2641.1

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_Taxid=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.F., Plunkett G. III, Plnick A., Peters B.T., Portland V.,

Piley M., Collado-Vides J., Glasner J.D., Fode N.W., Mayhew G.F.,

Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K12."

RL Science 277:1453-1474(1997).

RN [2]

RP IDENTIFICATION, AND CONCEPTUAL TRANSLATION.

RA Rudd K.E.;

RL Unpublished observations (JAN-1999).

CC -1- SIMILARITY: STRONG, TO E COLI YPFF

CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT HAD TO BE

SUPPRESSED IN POSITIONS 9 AND 11.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF000349; NOT_ANNOTATED_CDS.

DR EcoGene; EG13309; YPJ1.

RX Hypothetical protein; Complete proteome;

FW SEQUENCE 90 AA; 10070 MW; E103E2F6C2E8B697 CPO64;

SO SEQUENCE

Query Match 100.0%; Score 31; DB 1; Length 90;

Best Local Similarity 50.0%; Pred No 1 8a+02;

Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6

DB 61 WRWVNF 66

RESULT 7

YG50_MYCPN

ID_YG50_MYCPN STANDARD; PRT; 101 AA.

AC P75147;

DT 16-OCT-2001 (Rel 40, Created)

DT 16-OCT-2001 (Rel 40, Last sequence update)

DT 15-JUN-2002 (Rel 41, Last annotation update)

DE Hypothetical lipoprotein YG50 precursor (E09_01101).

GN YG50 OR MF192

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_Taxid=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129.

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelreich R., Hilbert H., Plogers H., Pirkl E., Li B.C.,

Herrmann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

[illegible]

Query Match	100.0%	Score 317	DB 1	Length 1017		
Best local similarity	100.0%	174.0	100.0%			
Matches	31	Conservation	2	Min+Max	0	100.0%

QY	1	WXXXWF	6
ED	95	WTNWKF	100

```

RESULT 8
NTSM_HUMAN
ID NTSM_HUMAN STANDARD PT 105 AA
AC 065178;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NADH-ubiquinone dehydrogenase, cytochrome b5L complex (CI-ACGGC)
DE (EC 1.6.5.3) (EC 1.6.99.3) (Complex I-ACGGC) (CI-ACGGC)
GN NDUFB2.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Mammalia; Primates; Catarrhini; Hominoidea; Homo
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE 20443367, PubMed 1042155,
RA Joffen T, W, Gillet F, Van den Heuvel A, Zampieri M,
RA Desbats C A F, Gerecht P F, Tribel V W F, Gellera A W,
RA "Genome of eight human and baboon cells: a comparative
RA comparative study of human and baboon genomes and the evolution of
RA biochem. Biophys. Res. Comm. 1994; 204:100000"
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Blood,
RX MEDLINE 20443367, PubMed 1042155,
RA Zhang Q, H, Yu W, Wu X, Y, Fan G-X, Zhao M, Zhao G-Q, Fu G-Y,
RA Shen Y, Fan H-Y, Lu G, Zhou M, Xu X-S, Han G-Q, Chang T-W,
RA Tao J, Huang Q-H, Zhou J, Hu G-X, Gu J, Chen S-J, Chen Z-F,
RA "Cloning and functional analysis of cDNA with full-length cDNA for
RA 300 protein-ubiquitin protein complex in 2004. 1042155"
RT stem/progenitor cells"
RI Genome Res. 10:1040-1060(2004)
RN [3]
RP SEQUENCE FROM N.A
RA TISSUE=Lung;
RA Straussberg R.J
RA Submitted (Apr 2004) to the EMBL/GenBank/DDP database

```

00 -1- REMOVED FROM THE EMBL DATABASE FOR THE REASON THAT THE SEQUENCE
01 HAVING THE SAME SEQUENCE AS THE EMBL ENTRY FOR THE ENZYME IS BELIEVED
02 TO BE UNRELIABLE.
03
04 -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
05
06 -1- CATALYTIC ACTIVITY WITH OTHER SUBSTRATES: NADH + ubiquinol
07
08 -1- SUBUNIT COMPLEX: A COMPLEX OF ABOUT 10 DIFFERENT SUBUNITS.
09
10 -1- SUBCELLULAR LOCATION: Mitochondrion (inner membrane, matrix side).
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
8

*seq1	Match	seq2	Match	seq3	Match	seq4	Match	seq5	Match	seq6	Match	seq7	Match	seq8	Match	seq9	Match	seq10	Match	seq11	Match	seq12	Match	seq13	Match	seq14	Match	seq15	Match	seq16	Match	seq17	Match	seq18	Match	seq19	Match	seq20	Match	seq21	Match	seq22	Match	seq23	Match	seq24	Match	seq25	Match	seq26	Match	seq27	Match	seq28	Match	seq29	Match	seq30	Match	seq31	Match	seq32	Match	seq33	Match	seq34	Match	seq35	Match	seq36	Match	seq37	Match	seq38	Match	seq39	Match	seq40	Match	seq41	Match	seq42	Match	seq43	Match	seq44	Match	seq45	Match	seq46	Match	seq47	Match	seq48	Match	seq49	Match	seq50	Match	seq51	Match	seq52	Match	seq53	Match	seq54	Match	seq55	Match	seq56	Match	seq57	Match	seq58	Match	seq59	Match	seq60	Match	seq61	Match	seq62	Match	seq63	Match	seq64	Match	seq65	Match	seq66	Match	seq67	Match	seq68	Match	seq69	Match	seq70	Match	seq71	Match	seq72	Match	seq73	Match	seq74	Match	seq75	Match	seq76	Match	seq77	Match	seq78	Match	seq79	Match	seq80	Match	seq81	Match	seq82	Match	seq83	Match	seq84	Match	seq85	Match	seq86	Match	seq87	Match	seq88	Match	seq89	Match	seq90	Match	seq91	Match	seq92	Match	seq93	Match	seq94	Match	seq95	Match	seq96	Match	seq97	Match	seq98	Match	seq99	Match	seq100	Match	seq101	Match	seq102	Match	seq103	Match	seq104	Match	seq105	Match	seq106	Match	seq107	Match	seq108	Match	seq109	Match	seq110	Match	seq111	Match	seq112	Match	seq113	Match	seq114	Match	seq115	Match	seq116	Match	seq117	Match	seq118	Match	seq119	Match	seq120	Match	seq121	Match	seq122	Match	seq123	Match	seq124	Match	seq125	Match	seq126	Match	seq127	Match	seq128	Match	seq129	Match	seq130	Match	seq131	Match	seq132	Match	seq133	Match	seq134	Match	seq135	Match	seq136	Match	seq137	Match	seq138	Match	seq139	Match	seq140	Match	seq141	Match	seq142	Match	seq143	Match	seq144	Match	seq145	Match	seq146	Match	seq147	Match	seq148	Match	seq149	Match	seq150	Match	seq151	Match	seq152	Match	seq153	Match	seq154	Match	seq155	Match	seq156	Match	seq157	Match	seq158	Match	seq159	Match	seq160	Match	seq161	Match	seq162	Match	seq163	Match	seq164	Match	seq165	Match	seq166	Match	seq167	Match	seq168	Match	seq169	Match	seq170	Match	seq171	Match	seq172	Match	seq173	Match	seq174	Match	seq175	Match	seq176	Match	seq177	Match	seq178	Match	seq179	Match	seq180	Match	seq181	Match	seq182	Match	seq183	Match	seq184	Match	seq185	Match	seq186	Match	seq187	Match	seq188	Match	seq189	Match	seq190	Match	seq191	Match	seq192	Match	seq193	Match	seq194	Match	seq195	Match	seq196	Match	seq197	Match	seq198	Match	seq199	Match	seq200	Match	seq201	Match	seq202	Match	seq203	Match	seq204	Match	seq205	Match	seq206	Match	seq207	Match	seq208	Match	seq209	Match	seq210	Match	seq211	Match	seq212	Match	seq213	Match	seq214	Match	seq215	Match	seq216	Match	seq217	Match	seq218	Match	seq219	Match	seq220	Match	seq221	Match	seq222	Match	seq223	Match	seq224	Match	seq225	Match	seq226	Match	seq227	Match	seq228	Match	seq229	Match	seq230	Match	seq231	Match	seq232	Match	seq233	Match	seq234	Match	seq235	Match	seq236	Match	seq237	Match	seq238	Match	seq239	Match	seq240	Match	seq241	Match	seq242	Match	seq243	Match	seq244	Match	seq245	Match	seq246	Match	seq247	Match	seq248	Match	seq249	Match	seq250	Match	seq251	Match	seq252	Match	seq253	Match	seq254	Match	seq255	Match	seq256	Match	seq257	Match	seq258	Match	seq259	Match	seq260	Match	seq261	Match	seq262	Match	seq263	Match	seq264	Match	seq265	Match	seq266	Match	seq267	Match	seq268	Match	seq269	Match	seq270	Match	seq271	Match	seq272	Match	seq273	Match	seq274	Match	seq275	Match	seq276	Match	seq277	Match	seq278	Match	seq279	Match	seq280	Match</
-------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	---------

QY	1	WXXWXF	6
		:: :	
Db	68	WILWRF	73

```

RESULT 9
YQCC_HAEIN STANDARD;     PRT;   106 AA.
ID YQCC_HAEIN
AC Q87157; O05061;
DT 15-JUL-1998 (Rel. 36, Created)
ET 15-JUL-1998 (Rel. 36, Last sequence update)
OT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical protein H1436.
GN H1436.
OS Haemophilus influenzae.
OC Proteobacteria, Gammaproteobacteria, Enterobacteriaceae;
OX NCBI_TaxID=727;
RN [1]
RZ SEQUENCE FROM N.A.
RA STRAIN K1 / KWID / ATCC 35067;
RA MEDLINE 3655063, PMID 1547807;
RA FINDERSCAN FT, March 1998, WU, China, Liang E.A., Kirkness E.F.,
RA Hollibaugh A.R., East CE, Smith D.E., Dougherty EA, Holland C.M.,
RA Peterson J., Liu Z., Fitzhugh M., Finkbeiner S., Tsayne D.,
RA Scott R., Shiley B., Liu L., Glaser A., Kelley G.M.,
RA Weidman J.E., Phillips C.A., Spriggs T., Hedblom F., Cotton M.D.,
RA Ulteback I.R., Hanna M.C., Nguyen D.T., Souder D.M., Branton R.C.,
RA Five L.D., Filchman J.L., Fuhsmann J.L., Georgagen N.S.M.,
RA Green C.L., McDonald L.W., Small R.V., Finck C.M., Smith H.O.,
RA Venter J.C.;
RA While genome random sequencing and assembly of Haemophilus influenzae
RT Rd.w;
RC Colides 163,436 Size(1975).
RI SIMILARITY TO THE N TERMINAL OF HAEMOPHILUS EXOGENOUS RESONATION
CC REGION ORPH. THE C TERMINAL PART IS COLINEAR WITH YQCC.
CC -1 SIMILARITY STRING: ID K0001 YQCC.
CC -----
CC This Swiss PROT entry is copyrighted. It is released through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

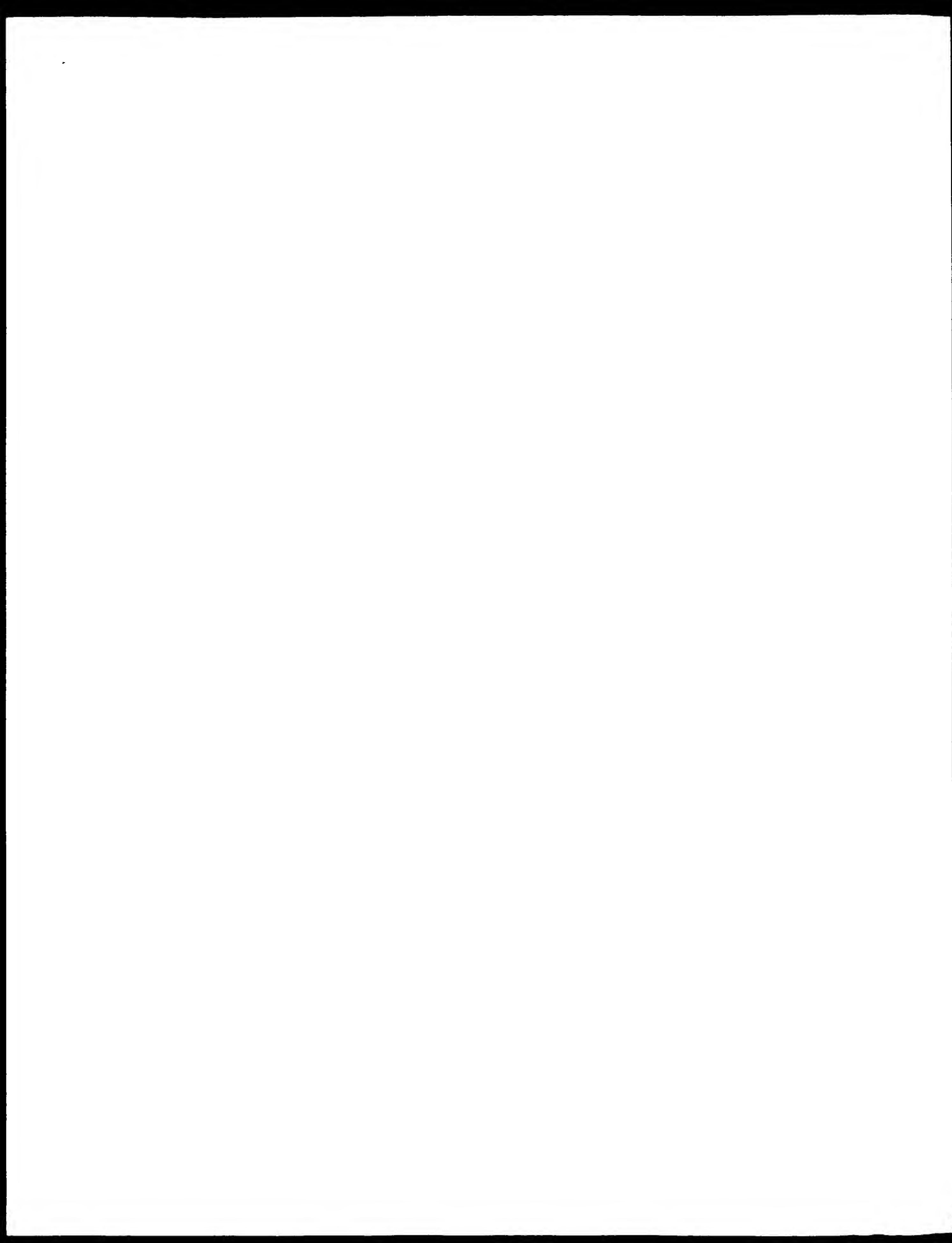

CC (MEMO):
CC -1- SUBCELLULAR LOCATION: Integral membrane protein
CC
CC This entry represents a protein that is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not profit institutions as long as the protein is in a way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch>
CC or send an email to license@isb-sib.ch
CC
CC EMBL: L29436; AAA98244.1; -
CC PIR: D29504; P29504
CC Transport: Mercury; resistance; Transmembrane; Mercury; Plasmid.
CC TRANSMEM 28 48 POTENTIAL.
CC TRANSMEM 71 91 POTENTIAL.
CC METAL 47 47 HG(2+) (POTENTIAL).
CC METAL 48 48 HG(2+) (POTENTIAL).
CC METAL 114 114 HG(2+) (POTENTIAL).
CC METAL 115 115 HG(2+) (POTENTIAL).
CC SEQUENCE 129 AA, 11102 MW, 23974EDB4A6A8FE91 CRO64;
Query Match 100.0%, Score 31, E 1, Length 129,
Best Local Similarity 50.0%, Freq. No. 2.4e+02;
Matches 3, Conservative 1, Mismatches 0, Indels 0, Gaps 0;
CY 1 WXXWXF 6
DB 28 WSWWAF 32
RESULT 13
K038_HUMAN
ID K038_HUMAN STANDARD, FRT, 129 AA.
AC P04207;
DI 20-MAR-1987 (Rel. 04, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DI 15-DEC-1999 (Rel. 39, Last annotation update)
DE 13 kappa chain V III region CIL precursor (P04207) [14471]
OS Homo sapiens (Human)
CC Eukaryotic, Mammal, Primate, Catarrhini, Hominidae, Homininae,
CC Mammalia, Eumetazoa, Chordata, Vertebrata, Euteleostomi,
CC NCBI_TaxID=9606;
CC
CC RN
CC
CC SEQUENCE FROM N.A.
RX MEDLINE-86177570; PubMed-3083417;
RA Goldstein R., Carson D.A.,
RT "Cloning and sequence determination of a human kappa chain V III
RT light-chain gene"
RD Proc. Natl. Acad. Sci. U.S.A. 83:2198 (1986)
CC
CC This SwissProt entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not profit institutions as long as the content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch>
CC or send an email to license@isb-sib.ch
CC
CC EMBL: M2240; AA583922.1; -
CC PIR: A01898; K3H0CL.
CC HSSP; P80362; IWT.
CC InterPro: IPR003596; I4_MN0.
CC InterPro: IPR003596; I4_MN0.
CC Pfam: PF00047; I4_1.
CC SMART; SM00406; IGV; 1.
CC Immunoglobulin V region; Signal.
CC SIGNAL 1 20
CC CHAIN 1 129 IG KAPPA CHAIN V-III REGION CIL.
CC DOMAIN 21 43 FRAMEWORK-1.
CC DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
CC DOMAIN 55 69 FRAMEWORK-2.

FT DOMAIN 70 76 COMPLEMENTARITY DETERMINING-2.
FT DOMAIN 77 106 FRAMEWORK-1.
FT DOMAIN 107 116 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 117 123 THE C-TERMINUS.
FT DISULFID 43 106 BY SIMILARITY.
FT NON-TER 129 129
FT SEQUENCE 129 AA, 14102 MW, 23974EDB4A6A8FE91 CRO64;
Query Match 100.0%, Score 31, E 1, Length 129,
Best Local Similarity 50.0%, Freq. No. 2.4e+02;
Matches 3, Conservative 1, Mismatches 0, Indels 0, Gaps 0;
CY 1 WXXWXF 6
DB 64 WTRWSF 69
RESULT 15
Y091_MYCN
ID Y091_MYCN STANDARD, FRT, 129 AA.
AC P38857;
DI 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DI 10-DEC-1999 (Rel. 40, Last annotation update)
DE 13 kappa chain V III region CIL precursor (P38857) [14471]
OS Homo sapiens (Human)
CC Eukaryotic, Mammal, Primate, Catarrhini, Hominidae, Homininae,
CC Mammalia, Eumetazoa, Chordata, Vertebrata, Euteleostomi,
CC NCBI_TaxID=9606;
CC
CC RN
CC
CC SEQUENCE FROM N.A.
RX MEDLINE-86177570; PubMed-3083417;
RA Goldstein R., Carson D.A.,
RT "Cloning and sequence determination of a human kappa chain V III
RT light-chain gene"
RD Proc. Natl. Acad. Sci. U.S.A. 83:2198 (1986)
CC
CC This SwissProt entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not profit institutions as long as the content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch>
CC or send an email to license@isb-sib.ch
CC
CC EMBL: D00027; AA583922.1; -
CC PIR: A01898; K3H0CL.
CC HSSP; P80362; IWT.
CC InterPro: IPR003596; I4_MN0.
CC InterPro: IPR003596; I4_MN0.
CC Pfam: PF00047; I4_1.
CC SMART; SM00406; IGV; 1.
CC Immunoglobulin V region; Signal.
CC SIGNAL 1 20
CC CHAIN 1 129 IG KAPPA CHAIN V-III REGION CIL.
CC DOMAIN 21 43 FRAMEWORK-1.
CC DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
CC DOMAIN 55 69 FRAMEWORK-2.

ID YA91_MYCPN STANDARD; PRT; 138 AA.
AC P75602;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN091 (P02_0f138).
GN MPN091 OR MP064.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN ATCC 29242 / M129,
RX MEDLINE-97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RI Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: ALMOST IDENTICAL TO M.PNEUMONIAE MPN413 AND MPN463.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000008; AAB95711.1; .
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 138 AA; 15763 MW; F52380D428F42709 CRC64;

Query Match 100.0%; Score 31; DB 1; Length 138;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 WXXWXF 6
Db 36 WSGWSF 41

Search completed: February 20, 2003, 10:48:02
Job time : 7.66667 secs



GenCure version 5.1.3
Copyright (c) 1991-2003 GIBCO-BRL

OM protein - protein search, using sw model

Run on February 20, 2003, 10:45:10 / Search time 23.333 Seconds
(without alignments)
52.994 Million cell updates/sec

Title: US-09-912-414-9
Perfect score: 31
Sequence: 1 WXXWXP 6

Scoring matrix: BLOSUM62EX
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206647115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 98
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriapi:*
- 17: sp_archaeapi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	31	100.0	28	6	Q62821 Bubalus bub
2	31	100.0	31	8	Q94NH2 Bittorina s
3	31	100.0	36	15	Q905D7 human immun
4	31	100.0	48	11	Q63365 ratirus ngrv
5	31	100.0	49	11	Q08615 mus musculu
6	31	100.0	50	4	Q9NRG0 Q9NRG0 homo sapien
7	31	100.0	59	16	Q9PH20 Q9PH20 xylella fas
8	31	100.0	61	16	Q85619 escherichia
9	31	100.0	64	4	Q8WVM9 Q8WVM9 homo sapien
10	31	100.0	70	2	Q9EXE4 Q9EXE4 lactobacteri
11	31	100.0	71	17	Q9HST8 Q9HST8 halobacteri
12	31	100.0	74	2	Q50168 Q50168 mycobacteri
13	31	100.0	76	16	Q68793 Q68793 yersinia pe
14	31	100.0	77	2	Q9S4W3 Q9S4W3 escherichia
15	31	100.0	79	2	Q9WTF7 Q9WTF7 escherichia
16	31	100.0	79	16	Q9JMR3 Q9JMR3 escherichia

17	31	100.0	80	16	Q93GPK Q93GPK salmonella
18	31	100.0	83	16	Q8UEM9 Q8UEM9 agrobacteri
19	31	100.0	86	2	Q9WTF7 Q9WTF7 escherichia
20	31	100.0	86	2	Q923FA Q923FA escherichia
21	31	100.0	86	10	Q8WIM2 Q8WIM2 solanum cha
22	31	100.0	87	16	Q93GPK Q93GPK solanum pin
23	31	100.0	89	10	Q8WIL9 Q8WIL9 solanum pin
24	31	100.0	89	10	Q8WIL9 Q8WIL9 solanum pin
25	31	100.0	92	10	Q8WIM0 Q8WIM0 solanum bul
26	31	100.0	92	12	Q90990 Q90990 turkey rin
27	31	100.0	93	10	Q8WIM1 Q8WIM1 solanum bul
28	31	100.0	93	16	Q9UEW1 Q9UEW1 escherichia
29	31	100.0	94	4	Q9WTF7 Q9WTF7 escherichia
30	31	100.0	96	4	Q96QD1 Q96QD1 homo sapien
31	31	100.0	97	16	Q9KSV6 Q9KSV6 bacillus ba
32	31	100.0	99	16	Q94AN0 Q94AN0 arabidopsis
33	31	100.0	99	16	Q94AN0 Q94AN0 arabidopsis
34	31	100.0	101	5	Q95RS1 Q95RS1 drosophila
35	31	100.0	104	10	Q9X162 Q9X162 tritium ar
36	31	100.0	103	16	Q9KTF5 Q9KTF5 vibrio chol
37	31	100.0	103	16	Q9XX10 Q9XX10 ralscena s
38	31	100.0	105	11	Q9CPU2 Q9CPU2 mus musculu
39	31	100.0	106	16	Q9CNE4 Q9CNE4 pasteurella
40	31	100.0	106	16	Q9PSE6 Q9PSE6 xylella fas
41	31	100.0	106	16	Q9K1V7 Q9K1V7 chlamydia p
42	31	100.0	106	16	Q92FT9 Q92FT9 chlamydia p
43	31	100.0	107	2	Q9665R Q9665R arabidopsis
44	31	100.0	108	16	Q9WTF7 Q9WTF7 escherichia
45	31	100.0	108	16	Q9WTF7 Q9WTF7 escherichia

ALIGNMENTS

RESULT 1
ID O62821 PRELIMINARY, FRT, 28 AA.
AC O62821;
DT 01-AUG-1998 (TREMBLrel, 07, Created)
DT 01-AUG-1998 (TREMBLrel, 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DE Lysozyme (Fragment).
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID-89462;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RA Das P., Tiwari G., Garg L.C.;
RT "Characterisation of Lysozyme encoding gene promoter in Bubalus bubalis";
RT Submitted (MAP-1998) to the EMBL/GenBank/DBJ databases.
PI, EMBL, AJ225012; CAA12342.1; --
DR EMBL, AJ225012; CAA12342.1; --
FT NON TER 28
FT 28
SQ SEQUENCE 28 AA; 3160 MW; RQGF666AAT7TDB290 Q9064;

Query Match 100.0%, Score 31, 16 E, Length 28;
Post local similarity 50.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWXP 6
DB 10 WSWWTF 15

RESULT 2

ID Q94NH2 PRELIMINARY, PRT, 31 AA.
AC Q94NH2;
DT 01-DEC-2001 (TREMBLrel, 19, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)

DT 01-OCT-2000 (TREMBlrel 15, Created)
DT 01-OCT-2000 (TREMBlrel 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel 19, Last annotation update)
DE FAM4A1 (Fragment).
GN FAM4A1.
OS Homo sapiens (Human)
OC Eukaryota, Metazoa, Chordata, Vertebrata, Eumetazoa, Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.
OC NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N A
RX MEDLINE=20349222; PubMed=10889047,
RA Vincent J R, Hethcote J A, Gurling H M D, Bolton P F, Roberts W,
RA Scherer S W;
RT "Identification of a novel gene on chromosome 7q31 that is interrupted
RT by a translocation breakpoint in an autistic individual."
RL Am. J. Hum. Genet. 67:510-514(2001).
DR EMBL: AF234886; AAF85949.1;
FT NON_TER 50
SQ SEQUENCE 50 AA; 5441 MW; 545851F74DDP23 CPG64;

Query Match 100.0%; Score 31; DB 4; Length 50;
Best Local Similarity 50.0%; Pred No 50+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXWXF 6
DB 24 WTWFF 29

RESULT 7
Q9PH20
ID Q9PH20 PRELIMINARY; PRT; 50 AA
AC Q9PH20;
DT 01-OCT-2000 (TREMBlrel 15, Created)
DT 01-OCT-2000 (TREMBlrel 15, Last sequence update)
DT 01-MAY-2002 (TREMBlrel 21, Last annotation update)
DE Hypothetical protein Xf0126.
GN Xf0126.
OS Xylella fastidiosa
OC Prokaryota, Bacteria, Gammaproteobacteria, Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN 11
RP SEQUENCE FROM N A
RX STRAIN=9A5C;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A J G, Reinach F C, Arruda P, Abreu F A, Agencio M,
RA Alvaranga P, Alves L M C, Araya T E, Bala G S, Baptista C S,
RA Barros M H, Bonatoni F B, Bordin S, Bove J M, Briones M P S,
RA Buena M P P, Camargo A A, Camargo L E A, Carraro B M, Carter H,
RA Colares R R, Colares C, Costa F F, Costa M C P, Costa Neto C M,
RA Coutinho L L, Cristofani M, Dias Neto F, Docena C, El Dorry H,
RA Facincani A P, Ferreira A T S, Ferreira V C A, Ferro J A,
RA Fraga T S, Franca S C, Franco M C, Frohne M, Fujita L R,
RA Garnier M, Goldman G H, Goldman M H S, Gomes S L, Gruber A,
RA Ho P L, Hobeisel J D, Junqueira M B, Kempf E L, Kitajima T P,
RA Krieger J E, Kurama E E, Laigret F, Lambais M P, Leite L C C,
RA Lemos F G M, Lemos M V F, Lopes S A, Lopes C P, Machado J A,
RA Machado M A, Madeira A M B N, Madeira H M F, Marino C L,
RA Marques M V, Martins E A L, Martins E M F, Matsukuma A Y,
RA Menck C F M, Miranda E C, Miyaki C Y, Monteiro-Vitorello C B,
RA Moon D H, Nagai M A, Nascimento A L T O, Netto L F C,
RA Nhani A, Jr, Nobrega F G, Nunes L P, Oliveira M A,
RA de Oliveira M C, de Oliveira P C, Palmieri P A, Paris A,
RA Peixoto R P, Pereira G A G, Pereira H A Jr, Pasquero J B,
RA Quadrio R R, Roberto P G, Rodrigues V, de Rosa A J M,
RA de Rosa V R, Jr, de Sa P G, Santelli P V, Sawasaki H E,
RA da Silva A C R, da Silva A M, da Silva F R, Silva W A Jr,
RA da Silva J F, Silvestri M L Z, Siqueira W D, de Souza A A,
RA de Souza A P, Toranzo M F, Truffi P, Tsai S M, Tsubaki M H,
RA Vallada H, Van Sluys M A, Verjovski-Almeida S, Vettore A B,
RA Zago M A, Zatz M, Zeldanis J, Zetubal J C;

PT "The genome sequence of the plant pathogen Xylella fastidiosa."
PL Nature 406:151-159(2000).
DR EMBL: AE002846; AAF8949.1;
DR InterPro: IPR001230; Preyyl_site
DR PROSITE: PS00294; PRENYLATION, UNKNOWN_1.
RW Hypothetical protein; Complete Proteome
SQ SEQUENCE 50 AA; 5441 MW; 545851F74DDP23 CPG64;

Query Match 100.0%; Score 31; DB 16; Length 50;
Best Local Similarity 50.0%; Pred No 50+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXWXF 6
DB 54 WPCWLF 59

RESULT 8
O85619
ID O85619 PRELIMINARY; PRT; 61 AA.
AC O85619;
DT 01-NOV-1998 (TREMBlrel 08, Created)
DT 01-NOV-1998 (TREMBlrel 08, Last sequence update)
DT 01-MAY-2002 (TREMBlrel 20, Last annotation update)
DE L0012 (Hypothetical 5.5 kDa protein).
GN Z5095 OR ECS4544.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN 11

RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=EDL933;
RX MEDLINE=94429885; PubMed=9477246;
RA Perna N T, Mayhew G F, Postel G, Elliott G, Donnenberg M S,
RA Kaper J B, Plautner F P;
RT "Molecular evolution of a pathogenicity island from enterohemorrhagic
RT Escherichia coli O157:H7."
PL Infect Immun 66:3810-3817(1998).
RN 121
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700922;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N T, Plunkett G III, Burland V, Mau B, Glasner J D,
RA Rose D J, Mayhew G F, Evans P S, Gregor J, Kirkpatrick H A,
RA Postel G, Hackett J, Klink S, Boutin A, Shao Y, Miller L,
RA Grobeck B J, Davis N W, Lim A, Dimalanta E T, Polomous K,
RA Apodaca C, Anantharaman T S, Lin J, Yen G, Schwartz P C,
RA Welch P A, Blattner F P;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7."
PL Nature 409:520 523(2001)
RN 123
RP SEQUENCE FROM N A
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EHEC-ecv0040;
RX MEDLINE=21156031; PubMed=11268796;
RA Hayashi T, Makino K, Ohnishi M, Kurokawa Y, Teshi Y, Yokoyama K,
RA Han C G, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T,
RA Iida T, Takami H, Honda T, Sasakawa C, Ogisawara N, Yasunaga T,
RA Kuhara S, Shiba T, Hattori M, Shinagawa H;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
PL DNA Res. 8:11-22(2001).
DR EMBL: AF071034; AAC31491.1;
DR EMBL: AE005594, AAG58810.1, ALT_INIT.
DR EMBL: AP002566, BAB37967.1;
RW Hypothetical protein; Complete Proteome
SQ SEQUENCE 61 AA; 7286 MW; 5B7F4811F9D58CD3 CPG64;

Query Match 100.0%; Score 31; DB 16; Length 61;
Best Local Similarity 50.0%; Pred No 50+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

068793
ID 068793 PRELIMINARY; PRT; 76 AA.
AC 068793;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein Y1031.
GN Y1091 OR YPMT1.600.
OS Yersinia pestis.
OG Plasmid PMT1 (PMT-1).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae,
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5;
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott T., McCready P., Skowronski F., Carnes D.,
RA Kobayashi A., Brubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Yersinia
pestis.";
RL J. Bacteriol. 180:5193-5202(1998)
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM10+;
RX MEDLINE=99043898; PubMed=9826348;
RA Lindler T.F., Plane G.V., Buland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
KIM5 plasmid encoding murine toxin and capsular antigen."
RL Infect. Immun. 66:5731-5742(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BICVAR ORIENTALIS; PLASMID=PMT1 (PMT 1);
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill T., Wren R.W., Thomson N.P., Titchall P.W., Holden M.T.G.,
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies P.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Harty J.S., Jagels K., Jarvis D.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skellern J., Stevens K., Whitehead S., Barrall R.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague",
RL Nature 413:523-527(2001).
DR EMBL; AF053947; AAC13243.1; -;
DR EMBL; AF074611; AAC82749.1; -;
DR EMBL; AL117211; CAB55242.1; -;
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 76 AA; 8407 MW; 7FF44917744946FB CPG64;

Query March 100.0%; Score 31; PP 16; Length 76;
Best Local Similarity 50.0%; Pred. No. 7; Neglog;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXF 6
Db 50 WRTWNF 55

RESULT 14
Q9S4W3 PRELIMINARY; PRT; 77 AA.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE Hypothetical 9.1 kDa protein.
OS Escherichia coli.
OG Plasmid F.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

FX MEDLINE=99296678; PubMed=10366527;
RA Manwaring N.P., Skurray R.A., Firth N.;
RT "Nucleotide sequence of the F plasmid leading region.";
RL Plasmid 41:219-225(1999).
DR EMBL; AF106329; AAB47189.1; -;
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 77 AA; 9083 MW; 3E43BF0257B377A CPG64;

Query March 100.0%; Score 31; PP 2; Length 77;
Best Local Similarity 50.0%; Pred. No. 7.4e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXF 6
Db 53 WRTWNF 58

RESULT 15
Q9WTF7 PRELIMINARY; PRT; 79 AA.
AC Q9WTF7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Ydea protein.
OS Ydea.
OS Escherichia coli.
OG Plasmid P100.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Samped G., Mizobuchi K.;
RT "Organization and diversification of plasmid genomes: complete
PT nucleotide sequence of the p100 genome."
PL Submitted (MAY-1999) to the EMBL/Genbank/DBP databases.
DR EMBL; AP000342; BAA78827.1; -;
SQ SEQUENCE 79 AA; 9328 MW; 7176F48F7C70156F CPG64;

Query March 100.0%; Score 31; PP 2; Length 79;
Best Local Similarity 50.0%; Pred. No. 7.4e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXF 6
Db 55 WRTWNF 60

Search completed: February 20, 2003, 10:49:20
Job time: 05.333 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein protein search, using sw model

Run on: February 20, 2003, 11:21:07, Search time 29.133 seconds
(without alignment)
27.256 Million seqs updated/sec

Title: us-09-912-414-2
Perfect score: 6
Sequence: 1 WYRWHP 6

Scoring table: OLIGOX
Gapop 60 0, Gapext 60 0

Searched: 908470 seqs, 133250620 residues

Word size: 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: A_Geneseq_101002:*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	6	21 AAB01492	Peptide which bind
2	5	83.3	6	21 AAB01504	Peptide which bind
3	5	83.3	6	21 AAB01509	Peptide which bind
4	5	83.3	372	20 AAY08982	Human tPA receptor
5	5	83.3	372	20 AAY05493	Human PEP 24 prote
6	5	83.3	373	22 AAU00227	tPA receptor-relat
7	4	66.7	6	14 AAR37389	Peptide for treati
8	4	66.7	6	21 AAB01505	Peptide which bind
9	4	66.7	6	21 AAB01508	Peptide which bind
10	4	66.7	10	17 AAR86140	Anti-ELAM-1 bindin

11	4	66.7	10	17	AAR86145	Anti ELAM-1 bindin
12	4	66.7	10	17	AAR86146	Anti-ELAM-1 bindin
13	4	66.7	10	19	AAW63963	ELAM-1 peptide mim
14	4	66.7	10	19	AAW63964	ELAM-1 peptide mim
15	4	66.7	10	19	AAW63958	ELAM-1 peptide mim
16	4	66.7	12	14	AAW63519	ELAM-1 peptide mim
17	4	66.7	12	15	AAW65756	Random peptide #53
18	4	66.7	12	17	AAW65754	Random peptide #53
19	4	66.7	12	18	AAW65286	Antibody D32.39 ep
20	4	66.7	12	20	AAW60380	Antibody eluted af
21	4	66.7	14	23	ABE74383	Karyophilic peptid
22	4	66.7	27	22	AAQ13264	Human P-tyrosinid
23	4	66.7	21	22	ABE43615	Peptide #1121 enc
24	4	66.7	31	22	ABE36569	Protein #9568 enco
25	4	66.7	31	22	AAW64556	Human brain expres
26	4	66.7	31	22	AAW77359	Human bone marrow
27	4	66.7	31	22	AAW21282	Peptide #3716 enco
28	4	66.7	31	22	AAW37509	Peptide #11546 enc
29	4	66.7	31	23	ABE36373	Human peptide enco
30	4	66.7	35	22	AAW87677	Human immune/haema
31	4	66.7	37	22	AAQ13641	Human poly-peptide
32	4	66.7	38	22	ABE30966	Peptide #3617 enco
33	4	66.7	38	22	ABE36158	Peptide #3664 enco
34	4	66.7	38	22	ABE21544	Protein #3543 enco
35	4	66.7	38	22	AAW69327	Human bone marrow
36	4	66.7	38	22	AAQ12011	Human poly-peptide
37	4	66.7	38	22	AAW17157	Peptide #3591 enco
38	4	66.7	38	22	AAW12648	Peptide #3585 enco
39	4	66.7	38	22	AAW04851	Peptide #3583 enco
40	4	66.7	38	22	ABE38943	Human peptide enco
41	4	66.7	41	22	AAQ13303	Human poly-peptide
42	4	66.7	42	22	AAW12672	Human bone marrow
43	4	66.7	42	22	AAW13865	Peptide #351 enco
44	4	66.7	48	22	ABE43116	Drosophila melanog
45	4	66.7	48	22	AAQ11856	Human poly-peptide

ALIGNMENTS

RESULT 1
ID AAB01492 standard, peptide, 6 AA.
XX AAB01492;
XX 09 NOV 2000 (first entry)
XX Peptide which binds to transcription factor E2F: DNA binding domain.
XX DE
XX DNA binding; transcription factor, E2F, E2F-1, cell cycle, PP-1;
XX FW activation; transcription, apoptosis, proliferative disorder;
XX FW psoriasis; restenosis.
XX OS
XX Synthetic.
XX PN WC200044771-A1.
XX 03-AUG-2000.
XX PD
XX 26-JAN-2000, 2000W0 GB00227.
XX PF
XX 26-JAN-1999; 99GB-0001710.
XX PR
XX (PROT-) PROLIFIX LTD.
XX PI Mueller R, Kontermann RE, Montigiani S;
XX WPI; 2000-532806/48.
XX Peptides binding to the DNA binding domain of transcription factor E2F
XX and inhibiting cell cycle progression, useful for the treatment of
XX cancer

KW tumour cell invasion, neurotransmitter release, chemotaxis, cancer;
 KW drug design; tumour; anti coagulant; vascular restenosis, arthritis;
 KW anti-inflammatory; wound healing; tissue regeneration; blood coagulation;
 KW nerve regeneration; osteoporosis; bone regeneration; cosmetic; scarring;
 KW ischaemic heart disease; myocardial infarction; angina pectoris; stroke;
 KW cerebral infarction; cerebral sequelae; subarachnoid haemorrhage;
 KW benign prostatic hyperplasia; diabetes mellitus; retinopathy, neuropathy,
 KW nephropathy; diabetic retinopathy.
 XX
 OS Homo sapiens.
 XX
 PN WO9929887-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WO-US26262.
 XX
 PR 10-DEC-1997; 97US-0987943.
 XX
 PA (SYNA-) SYNAPTIC PHARM CORP.
 XX
 PI Bard JA;
 XX
 DR WPI; 1999-385620/32.
 DR N-PSDB; AAX78110
 XX
 PT Lysophosphatidic acid receptor sequence
 XX
 PS Claim 9; Fig 3A-B; 121pp; English.
 XX
 CC This invention describes a novel nucleic acid (I) encoding a human
 CC lysophosphatidic acid (LPA; 1-acyl-glycerol-3-phosphate) receptor. Probes
 CC and antibodies which bind to the LPA receptor can be used for detecting
 CC expression of a mammalian LPA receptor. The probe can be used for
 CC detecting a predisposition to a disorder associated with LPA receptor
 CC activity. The transgenic animals incorporating the LPA receptor can be
 CC used for determining the physiological effects of varying levels of
 CC activity of mammalian LPA receptors, or for identifying an (ant)agonist
 CC capable of alleviating a physical or behavioural abnormality associated
 CC with activity of LPA receptor. LPA receptors produce a wide range of
 CC responses such as cell proliferation, stimulation of neurite retraction,
 CC platelet aggregation, smooth muscle contraction, tumour cell invasion,
 CC neurotransmitter release, chloride efflux and chemotaxis. The human LPA
 CC receptor receptor can serve as a tool for the drug design of novel
 CC therapeutic agents for various indications, including cancer, tumour
 CC reduction or prevention, platelet aggregation (as an anti-coagulant),
 CC vascular restenosis, arthritis (as an anti-inflammatory), wound healing,
 CC tissue regeneration (preferably skin and nerve regeneration), blood
 CC coagulation, osteoporosis (bone regeneration), and cosmetic uses
 CC (preferably the prevention of abnormal growths or scarring or for
 CC augmentation). This receptor can additionally serve as a tool for the
 CC drug design of novel therapeutic agents for indications including
 CC ischaemic heart diseases (e.g. myocardial infarction and angina
 CC pectoris), stroke due to cerebrovascular disturbances (e.g. cerebral
 CC infarction and cerebral sequelae after subarachnoid haemorrhage), benign
 CC prostatic hyperplasia, complications of diabetes mellitus (e.g.
 CC retinopathy, neuropathy, and nephropathy), and diabetic retinopathy.
 XX
 SQ Sequence 372 AA;
 OY 2 VRWHF 6
 Db 98 VRWHF 102

AC AAY05493;
 XX
 DT 07-JUL-1999 (first entry)
 XX
 DE Human PSP-24 protein sequence.
 XX
 KW FPG-1, FPG-2, FPG-3, FPG 4, FPG 5, PSP 24, human, detection; therapy;
 KW inverse agonist; allosteric modulator; lysophosphatidic acid receptor;
 KW LPA signalling mediated disease; cellular apoptosis.
 XX
 OS Homo sapiens.
 XX
 PN WO9919513-A2.
 XX
 PD 22-APR-1999.
 XX
 PF 09-OCT-1998; 98WO-US21315.
 XX
 PR 10-OCT-1997; 97US-0061572.
 XX
 PA (LXRB-) LXR BIOTECHNOLOGY INC
 XX
 PI Erikson J, Goddard JG, Kiefer M;
 XX
 DR WPI; 1999-377658/23.
 DR N-PSDB; AAX36570.
 XX
 PT Identification of (ant)agonists of LPA receptor EDG-2 for, e.g.
 PT treating LPA signalling mediated diseases such as cellular apoptosis
 XX
 PS Claim 1; Page 58-60; 63pp; English.
 XX
 CC This sequence represents human PSP-24.
 CC The invention relates to methods of detecting (ant)agonist, inverse
 CC agonist or allosteric modulators of the lysophosphatidic acid receptors
 CC EDG-1, EDG-2, FPG-3, FPG-4, FPG-5, and PSP-24. The methods are used to
 CC identify (ant)agonists and allosteric modulator of the lysophosphatidic
 CC acid (LPA) EDG2 receptor, e.g. to treat LPA signalling mediated disease
 CC such cellular apoptosis.
 XX
 SQ Sequence 373 AA;
 OY 2 VRWHF 6
 Db 98 VRWHF 102

RESULT 6
 AAU00227
 ID AAU00227 standard; Protein; 373 AA.
 XX
 AC AAU00227;
 XX
 DT 11-MAY-2001 (first entry)
 XX
 DE LPA receptor-related amino acid sequence #1.
 XX
 KW LPA receptor, EDG-2; lysophosphatidic acid, phospholipid, tumour;
 KW cell signalling, MAP kinase; LPA modulator; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; neuron damage; apoptosis;
 KW ischaemic heart disease; viral; HIV; inflammatory bowel disease;
 KW organ transplant; mouse.
 XX
 OS Mus sp.
 XX
 PN WO200112838-A2.
 XX
 PD 22-FEB-2001.
 XX

[illegible]

XX	(G)WP/194480 M.
XX	
P1	Folio M;
XX	
PP	WPI; 1993-094304/11.
XX	
PT	New peptide for treatment of prevention of viral infection comprises specified sequences of amino acid(s) and analogs
PT	comprising sequence informationated
XX	
PS	Example; Page 5; 39pp; English.
XX	
CC	The (G)WP/194480 peptide is an example of a specific peptide of formula
CC	1) $\text{H}_2\text{N}-\text{A}(\text{CH}_2)_n-\text{NH}_2$, where A is a substituted amino acid residue, n=1-100
CC	and each P is H, an amino acid residue or a fatty acid residue.
CC	The peptide is useful for treating or preventing septic shock,
CC	taking with polymyxin B, penicillin G, tetracycline, streptomycin
CC	or other antibiotics, or in combination with blood or plasma or their derivatives or in
CC	diagnostic reagents to detect and identify bacteria in blood
CC	or urine, preparing pharmaceutical compositions of lipid A or
CC	lipid A derivatives (LPS), and for treating pertussis, bacterial
CC	conjunctivitis and HIV infection. The usual dose is 10-100
CC	mg/kg/day, given intravenously, it binds to the same sites as
CC	polymyxin B, i.e., it inhibits all the toxic effects of lipid A. It
CC	has no antibiotic activity, does not lyse erythrocytes, has no
CC	activity against tumor cells infected at confluence and is relatively unstable
CC	against proteases.
XX	
CC	Sequence 6 AA:
CC	
CC	Query Match 66.7%; Score 4; E-14; Length 6;
CC	Best local similarity 100.0%; Pred. No. 7.8e+05;
CC	Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps
CV	1 WVRW 4
DB	2 WVRW 5
RESULT 8	
AAB01505	
ID	AAB01505 standard; peptide; 6 AA.
XX	
AC	AAB01505;
XX	
DT	08-NOV-2000 (first entry)
DE	
XX	Peptide which binds to transcription factor: E2F 1 DNA binding domain.
XX	
KM	DNA binding; transcription factor; E2F, E2F-1, cell cycle; DP-1;
KW	activating transcription; apoptosis; proliferative disorder;
KW	psoriasis; restenosis.
XX	
C3	Cytoskeletal.
XX	
PN	MO2000044771-A1.
PD	
XX	03-AUG-2000.
XX	
PF	26-JAN-2000; 2000044771-A1.
XX	
HK	26 JAN-1999; 55GR-0001710.
XX	
PA	(PROL-) PROLIFIX LTD.
XX	
PI	Mueller R, Kontermann RE, Montigiani S;
XX	
DK	WPI; 2000-532806/48.
XX	
PT	Peptides binding to the DNA binding domain of transcription factor E2F
PT	and inhibiting cell cycle progression, useful for the treatment of
PT	cancer

AC AAR6145;
 XX
 DE 26 OCT-1996 (first entry)
 XX
 DE Anti-ELAM-1 binding peptide #117.
 XX
 KW Peptide mimetic; endothelial leukocyte adhesion molecule, ELAM, selectin, receptor; leukocyte; vascular wall; endothelium; extravasation;
 KW inflammation; stial; Lewis, cell surface glycoprotein; Hb60 cell.
 XX
 OS Synthetic.
 XX
 FH Key
 FT Modified-site 10
 FT /note= "contain amidated C-terminus"
 XX
 PN WC3331210-AL.
 XX
 PD 23 NOV-1995.
 XX
 PF 11-MAY-1995; 95WO-US06315.
 XX
 PK 11-MAY-1994; 94US-0241054.
 XX
 PA ABBV-1 ABBVMAX TECHNOLOGIES NV.
 XX
 PI Barrett RW, Cwiria SE, Dower WJ, Koller KT, Lee J,
 PI Martens CL, Ruhland-Fritsch B;
 XX
 WIPI; 1996-010687/01.
 XX
 DE New peptide(s) that bind to endothelial leukocyte adhesion molecule
 PT 1 - useful for treating inflammation and other E-selectin mediated
 PT diseases
 PT
 XX
 PE Disclosures; Page 17, 85pp, English.
 XX
 CC Peptide sequences are examples of peptides and their analogs
 CC that bind to endothelial leukocyte adhesion molecule (ELAM-1). This
 CC sequence is a member of the selectin family of receptors and is involved
 CC in binding of leukocytes to the vascular endothelial wall prior to
 CC extravasation of the leukocytes, e.g. to a site of inflammation.
 CC The peptides bind prefer to E-selectin but may also bind L- or
 CC P-selectin, and can be used to treat conditions mediated by E-selectin,
 CC e.g. inflammation and cell death. The peptides have strong affinity for the
 CC selectin receptors and inhibit the binding of the stial Lewis (SLe-x)
 CC part of cell surface glycoproteins to E-selectin. The peptide are
 CC small, generally less than 2 kD, have an ICD of up to 100 micro mole
 CC against binding of Hb60 cells to ELAM-1, have one or more peptide
 CC linkages replaced by chemoamine, phosphonate, thioether, thioester, amide,
 CC or NHCOR linkages where R = H or a lower alkyl and R' a lower alkyl.
 CC The peptides may also have substituted N- and C-termini e.g.
 CC amidated, N-benzyloxycarbonyl or N-tow alkyl opds.
 CC
 XX
 SQ Sequence 10 AA;
 XX
 Query Match 66.7%; Score 4; DB 17; Length 10;
 Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WVRW 4
 DB 6 WVRW 9
 XX
 RESULT 12
 AAR6146
 ID AAR6146 standard; peptide; 10 AA.
 XX
 AC AAR6146;
 XX
 DE 26 OCT-1995 (first entry)
 XX

DE Anti-ELAM-1 binding peptide #123.
 XX
 KW Peptide mimetic; endothelial leukocyte adhesion molecule, ELAM, selectin;
 KW receptor; leukocyte; vascular wall; endothelium; extravasation;
 KW inflammation; stial; Lewis, cell surface glycoprotein; Hb60 cell.
 XX
 OS Synthetic.
 XX
 FH Key
 FT Modified-site 10
 FT /note= "contain amidated C-terminus"
 XX
 PN WC3331210-AL.
 XX
 PD 23 NOV-1995.
 XX
 PF 11-MAY-1995; 95WO-US06315.
 XX
 PK 11-MAY-1994; 94US-0241054.
 XX
 PA ABBV-1 ABBVMAX TECHNOLOGIES NV.
 XX
 PI Barrett RW, Cwiria SE, Dower WJ, Koller KT, Lee J,
 PI Martens CL, Ruhland-Fritsch B;
 XX
 WIPI; 1996-010687/01.
 XX
 DE New peptide(s) that bind to endothelial leukocyte adhesion molecule
 PT 1 - useful for treating inflammation and other E-selectin mediated
 PT diseases
 PT
 XX
 PE Disclosures; Page 17, 85pp, English.
 XX
 CC Peptide sequences are examples of peptides and their analogs
 CC that bind to endothelial leukocyte adhesion molecule (ELAM-1). This
 CC sequence is a member of the selectin family of receptors and is involved
 CC in binding of leukocytes to the vascular endothelial wall prior to
 CC extravasation of the leukocytes, e.g. to a site of inflammation.
 CC The peptides bind prefer to E-selectin but may also bind L- or
 CC P-selectin, and can be used to treat conditions mediated by E-selectin,
 CC e.g. inflammation and cell death. The peptides have strong affinity for the
 CC selectin receptors and inhibit the binding of the stial Lewis (SLe-x)
 CC part of cell surface glycoproteins to E-selectin. The peptide are
 CC small, generally less than 2 kD, have an ICD of up to 100 micro mole
 CC against binding of Hb60 cells to ELAM-1, have one or more peptide
 CC linkages replaced by chemoamine, phosphonate, thioether, thioester, amide,
 CC or NHCOR linkages where R = H or a lower alkyl and R' a lower alkyl.
 CC The peptides may also have substituted N- and C-termini e.g.
 CC amidated, N-benzyloxycarbonyl or N-tow alkyl opds.
 CC
 XX
 SQ Sequence 10 AA;
 XX
 Query Match 66.7%; Score 4; DB 17; Length 10;
 Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WVRW 4
 DB 6 WVRW 9
 XX
 RESULT 13
 AAW61963
 ID AAW61963 standard; peptide; 10 AA.
 XX
 AC AAW61963;
 XX
 DE 02 OCT-1998 (first entry)
 XX
 DE ELAM-1 peptide mimetic #118.
 XX
 KW Endothelial leukocyte adhesion molecule 1, ELAM-1; inflammation;
 KW selectin; diagnosis; mimetic.
 XX
 OS Synthetic.
 XX
 FH Key
 FH Location/Qualifiers

FT Modified-site 10 /note="C-terminal Met is amidated"
 XX
 FT US5728802-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 12-MAY-1995; 92US-0439817.
 XX
 PR 12-MAY-1995; 95US-0439817.
 PR 06-MAY-1992; 92US-0881395.
 PR 05-MAY-1993; 93US-0057295.
 PR 11-MAY-1994; 94US-0241054.
 XX
 PA (AFFY-) AFFYMAX TECHNOLOGIES NV.
 XX
 PI Barrett RW, Cwiria SE, Dower WJ, Koller KJ, Lee J;
 PI Martens CL, Ruhland B;
 XX
 DP WPI; 1998-249882/22.
 XX
 PT Peptide(s) or their mimetic(s) that bind to E-selectin - useful for,
 PT e.g. treating conditions mediated by E-selectin such as inflammatory
 PT condition(s)
 XX
 PS Example 2; Column 93-94; 84pp; English.
 XX
 CC AAW63846-W64054 are peptides and peptide mimetics that bind selectins
 CC including endothelial leukocyte adhesion molecule 1 (ELAM-1) and can be
 CC used for blocking adhesion of leukocytes to the selectins. The peptides
 CC have applications for the treatment of conditions mediated by
 CC E-selectin, e.g. inflammatory conditions. They can also be used for
 CC diagnostic purposes, e.g. for identifying the vascular site of E-selectin
 CC in vivo or can be coupled to anti-inflammatory or other drugs.
 XX
 SQ Sequence 10 AA;
 Query Match 66.7%; Score 4; DR 19; Length 10;
 Best Local Similarity 100.0%; Pred No 9.6;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WVRW 4
 Db ||||
 6 WVRW 9
 RESULT 14
 AAW63964
 ID AAW63964 standard; peptide; 10 AA.
 XX
 AC AAW63964;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE ELAM-1 peptide mimetic #119.
 XX
 KW Endothelial leukocyte adhesion molecule 1; ELAM-1; inflammation;
 KW selectin; diagnosis; mimetic.
 XX
 OS Synthetic.
 XX
 PN US5728802-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 12-MAY-1995; 92US-0439817.
 XX
 PR 12-MAY-1995; 95US-0439817.
 PR 06-MAY-1992; 92US-0881395.
 PR 05-MAY-1993; 93US-0057295.
 PR 11-MAY-1994; 94US-0241054.
 XX
 PA (AFFY-) AFFYMAX TECHNOLOGIES NV.

XX
 PI Barrett RW, Cwiria SE, Dower WJ, Koller KJ, Lee J;
 PI Martens CL, Ruhland B;
 XX
 DP WPI; 1998-249882/22.
 XX
 PT Peptide(s) or their mimetic(s) that bind to E-selectin - useful for,
 PT e.g. treating conditions mediated by E-selectin such as inflammatory
 PT condition(s)
 XX
 PS Example 2; Column 93-94; 84pp; English.
 XX
 CC AAW63846-W64054 are peptides and peptide mimetics that bind selectins
 CC including endothelial leukocyte adhesion molecule 1 (ELAM-1) and can be
 CC used for blocking adhesion of leukocytes to the selectins. The peptides
 CC have applications for the treatment of conditions mediated by
 CC E-selectin, e.g. inflammatory conditions. They can also be used for
 CC diagnostic purposes, e.g. for identifying the vascular site of E-selectin
 CC in vivo or can be coupled to anti-inflammatory or other drugs.
 XX
 SQ Sequence 10 AA;
 Query Match 66.7%; Score 4; DR 19; Length 10;
 Best Local Similarity 100.0%; Pred No 9.6;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WVRW 4
 Db ||||
 6 WVRW 9
 RESULT 15
 AAW63958
 ID AAW63958 standard; peptide; 10 AA.
 XX
 AC AAW63958;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE ELAM-1 peptide mimetic #113.
 XX
 KW Endothelial leukocyte adhesion molecule 1; ELAM-1; inflammation;
 KW selectin; diagnosis; mimetic.
 XX
 OS Synthetic.
 XX
 PN US5728802-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 12-MAY-1995; 92US-0439817.
 XX
 PR 12-MAY-1995; 95US-0439817.
 PR 06-MAY-1992; 92US-0881395.
 PR 05-MAY-1993; 93US-0057295.
 PR 11-MAY-1994; 94US-0241054.
 XX
 PA (AFFY-) AFFYMAX TECHNOLOGIES NV.
 XX
 PI Barrett RW, Cwiria SE, Dower WJ, Koller KJ, Lee J;
 PI Martens CL, Ruhland B;
 XX
 DP WPI; 1998-249882/22.
 XX
 PT Peptide(s) or their mimetic(s) that bind to E-selectin - useful for,
 PT e.g. treating conditions mediated by E-selectin such as inflammatory
 PT condition(s)
 XX
 PS Example 2; Column 91-92, 84pp; English.
 XX
 CC AAW63846-W64054 are peptides and peptide mimetics that bind selectins
 CC including endothelial leukocyte adhesion molecule 1 (ELAM-1) and can be
 CC used for blocking adhesion of leukocytes to the selectins. The peptides

CC Have applications for the treatment of conditions mediated by
 CC E-selectin, e.g. inflammatory conditions. They can also be used for
 CC diagnostic purposes, e.g. for identifying the vascular site of E-selectin
 CC in vivo or can be coupled to anti-inflammatory or other drugs.

XX
 SQ Sequence 10 AA;

Query Match 66.7%; Score 4; DB 19; Length 10;
 Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WVRW 4
 Db 6 WVRW 9

Search completed: February 20, 2003, 11:26:31
 Job time: 30.3333 secs

GenCore version 5.1.3
Copyright 1993-2003 Computer Sciences

OW protein - protein search, using sw model

Run on: February 20, 2003, 11:26:38 / Search time 17.6667 seconds
(without alignments)
30.539 Million cell updates/sec

Title: US-09-912 414 2

Perfect score: 1 WVRWHP 6

Sequences: 1 WVRWHP 6

Sequences: 1 WVRWHP 6

Searched: 44830 seqs, 9991795 residues

Word size: 0

Total number of hits satisfying chosen parameters: 44830

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Pending_Patents_AA New:

1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.ppt.*
2: /cgn2_6/prodata/1/paa/US02_NEW_COMB.ppt.*
3: /cgn2_6/prodata/1/paa/US02_NEW_COMB.ppt.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.ppt.*
5: /cgn2_6/prodata/1/paa/US02_NEW_COMB.ppt.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.ppt.*
7: /cgn2_6/prodata/1/paa/US02_NEW_COMB.ppt.*

Pred No is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	83.3	372	US-10-225-567A-394	Sequence 394, App
2	4	66.7	20	PCT US02-34021-373	Sequence 373, App
3	4	66.7	20	PCT US02-34021-422	Sequence 422, App
4	4	66.7	38	US-10-057-498-11626	Sequence 11626, A
5	4	66.7	52	PCT US02-32727-5526	Sequence 5526, Ap
6	4	66.7	52	US-10-057-498-11626	Sequence 5526, Ap
7	4	66.7	52	US-10-057-498-11626	Sequence 5526, Ap
8	4	66.7	52	PCT US02-32727-7189	Sequence 7189, Ap
9	4	66.7	56	US-10-057-498-11626	Sequence 7189, Ap
10	4	66.7	56	US-10-057-498-11626	Sequence 7189, Ap
11	4	66.7	58	PCT US02-32727-11626	Sequence 11626, A
12	4	66.7	58	US-10-057-498-11626	Sequence 11626, A
13	4	66.7	64	US-10-057-498-11626	Sequence 11626, A
14	4	66.7	64	PCT US02-32727-24772	Sequence 24772, A
15	4	66.7	64	US-10-057-498-11626	Sequence 24772, A
16	4	66.7	64	US-10-057-498-11626	Sequence 24772, A
17	4	66.7	64	PCT US02-32727-24772	Sequence 24772, A
18	4	66.7	65	US-10-057-498-11626	Sequence 24772, A
19	4	66.7	65	US-10-057-498-11626	Sequence 24772, A
20	4	66.7	65	US-10-057-498-11626	Sequence 24772, A
21	4	66.7	66	PCT US02-32727-24772	Sequence 24772, A
22	4	66.7	66	US-10-057-498-11626	Sequence 24772, A
23	4	66.7	66	US-10-057-498-11626	Sequence 24772, A
24	4	66.7	67	US-10-057-498-11626	Sequence 24772, A
25	4	66.7	70	PCT US02-32727-24772	Sequence 24772, A
26	4	66.7	70	US-10-057-498-11626	Sequence 24772, A

27	4	66.7	70	PCT US02-32727-24772	Sequence 24772, A
28	4	66.7	74	PCT US02-32727-24772	Sequence 24772, A
29	4	66.7	74	US-10-057-498-11626	Sequence 24772, A
30	4	66.7	74	US-10-057-498-11626	Sequence 24772, A
31	4	66.7	79	PCT US02-32727-15180	Sequence 15180, A
32	4	66.7	79	US-10-057-498-11626	Sequence 15180, A
33	4	66.7	79	US-10-057-498-11626	Sequence 15180, A
34	4	66.7	82	US-10-057-498-11626	Sequence 15180, A
35	4	66.7	82	US-10-057-498-11626	Sequence 15180, A
36	4	66.7	102	PCT US02-32727-15180	Sequence 15180, A
37	4	66.7	102	US-10-057-498-11626	Sequence 15180, A
38	4	66.7	102	US-10-057-498-11626	Sequence 15180, A
39	4	66.7	107	PCT US02-32727-15180	Sequence 15180, A
40	4	66.7	107	US-10-057-498-11626	Sequence 15180, A
41	4	66.7	107	US-10-057-498-11626	Sequence 15180, A
42	4	66.7	113	PCT US02-32727-15180	Sequence 15180, A
43	4	66.7	113	US-10-057-498-11626	Sequence 15180, A
44	4	66.7	113	US-10-057-498-11626	Sequence 15180, A
45	4	66.7	122	US-10-057-498-11626	Sequence 15180, A

ALIGNMENTS

RESULT 1

US-10-225-567A-394

Sequence 394, Application US/10225567A

GENERAL INFORMATION:

APPLICANT: Lifespan Biosciences

APPLICANT: Brown, Joseph P.

APPLICANT: Koush, Christine L.

TITLE OF INVENTION: ANTISENSE PHOSPHONATE AND ANTISENSE RNA ANTISENSE COMPOUND PROPERTIES

FILE REFERENCE: 1920-4-4

CURRENT APPLICATION NUMBER: US/10225567A

CURRENT FILING DATE: 2001 12-19

PRIOR APPLICATION NUMBER: 60/257,144

PRIOR FILING DATE: 2000 12-19

NUMBER OF SEQ ID NOS: 2092

SOFTWARE: Patent version 3.1

SEQ ID NO 394

LENGTH: 372

TYPE: PRT

ORGANISM: Homo sapiens

US-10-225-567A-394

Query Match

US-10-225-567A-394 83.3% Score 5; DB 6; Length 372;

Post local similarity 100.0%, Pct 53;

Matches 5, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

Cy 2 VWRWHP 6
De 98 VWRWHP 102

RESULT 2

PCT US02-34021-373

Sequence 373, Application PC/TUS0234021

GENERAL INFORMATION:

APPLICANT: DGI Biotechnologies, Inc.

TITLE OF INVENTION: TAPET SPECIFIC SCREENING AND ITS USE FOR IDENTIFYING TAPET RIND

FILE REFERENCE: 2598-4009PC

CURRENT APPLICATION NUMBER: PCT/US02/34021

CURRENT FILING DATE: 2002 10-24

PRIOR APPLICATION NUMBER: 60/345,471

PRIOR FILING DATE: 2001 10-24

NUMBER OF SEQ ID NOS: 537

SOFTWARE: Patent version 3.1

SEQ ID NO 373

LENGTH: 20

TYPE: PRT

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: MISC FEATURE
OTHER INFORMATION: VEGFRI-20F 4 H1
PCT-US02-34021-373

Query Match 66.7%, Score 4, DB 1, Length 20,
Best Local Similarity 100.0%, Pred. No. 89,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0,

CY 2 VRWH 5
DB 11 VRWH 14

RESULT 3
PCT-US02-34021-422

Sequence 422, Application EC/TG3333333333
GENERAL INFORMATION:
APPLICANT: DGI Biotechnologies, Inc.
TITLE OF INVENTION: TARGET SPECIFIC SCREENING AND ITS USE FOR IDENTIFYING TARGET BIND
FILE REFERENCE: 2598-4009PG
CURRENT APPLICATION NUMBER: PCT/US02/34021
CURRENT FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: 62/747,171
PRIOR FILING DATE: 2001-10-24
NUMBER OF SEQ ID NOS: 537
SOFTWARE: PatentIn version 3.1
SEQ ID NO 422
LENGTH: 20
TYPE: PRT
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: MISC FEATURE
OTHER INFORMATION: VEGFR1-20F-A112
PCT-US02-34021-422

Query Match 66.7%, Score 4, DB 1, Length 20,
Best Local Similarity 100.0%, Pred. No. 88,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0,

CY 2 VRWH 5
DB 11 VRWH 14

RESULT 4
US-10-203-138A-13934

Sequence 13934, Application US/10203138A
GENERAL INFORMATION:
APPLICANT: Molecular Dynamics, Inc.
APPLICANT: Penn, Shalton G
APPLICANT: Fied, David R
APPLICANT: Hanzel, David K
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENE DERIVED SINGLE EXON NUCLEOTIDE ACID PROBE CONSTRUCT FOR
TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN BT 474
FILE REFERENCE: PB 0004 WO 8
CURRENT APPLICATION NUMBER: US/10/203,138A
CURRENT FILING DATE: 2002-08-02
PRIOR APPLICATION NUMBER: US 60/160,312
PRIOR FILING DATE: 04 February 1999
PRIOR APPLICATION NUMBER: 00/111,114
PRIOR FILING DATE: 26 May 2001
PRIOR APPLICATION NUMBER: 00/111,114
PRIOR FILING DATE: 03 August 2000
PRIOR APPLICATION NUMBER: 00/111,114
PRIOR FILING DATE: 03 September 2000
PRIOR APPLICATION NUMBER: 00/111,114
PRIOR FILING DATE: 27 September 2000
PRIOR APPLICATION NUMBER: 00/111,114
PRIOR FILING DATE: 21 September 2000
PRIOR APPLICATION NUMBER: 00/111,114
PRIOR FILING DATE: 30 June 2000
NUMBER OF SEQ ID NOS: 15438

SOFTWARE: Molecular Dynamics Software Distilling Engine
SEQ ID NO 13934
LENGTH: 38

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP 10 AC012143.1
FEATURE:
OTHER INFORMATION: EXTRACTED IN BT474, SIGNAL = 1.7
US-10-203-138A-13934

Query Match 66.7%, Score 4, DB 6, Length 38,
Best Local Similarity 100.0%, Pred. No. 1,4e+02,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0,

CY 2 VRWH 5
DB 1 VRWH 4

RESULT 5
PCT-US02-32727-5526

Sequence 5526, Application EC/TG3333333333
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Feisling, David
APPLICANT: Bhatia, Ajay
APPLICANT: Malsomeva, Tean, Elizabeth
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Shiyuan
APPLICANT: Lodes, Michael
APPLICANT: Jones, Robert
APPLICANT: Benson, Darin
APPLICANT: Bailey, Patrick
APPLICANT: Barth, Brenda
APPLICANT: Tegliss, John
TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
FILE REFERENCE: 210121 514C1
CURRENT APPLICATION NUMBER: PCT/US02/32727
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 30992
SEQ ID NO 5526
LENGTH: 52
TYPE: PRT
ORGANISM: Propionibacterium
PCT-US02-32727-5526

Query Match 66.7%, Score 4, DB 1, Length 52,
Best Local Similarity 100.0%, Pred. No. 1,4e+02,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0,

CY 2 VRWH 5
DB 26 VRWH 29

RESULT 6
US-10-203-138A-13934

Sequence 13934, Application US/10203138A
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Feisling, David
APPLICANT: Bhatia, Ajay
APPLICANT: Malsomeva, Tean, Elizabeth
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Shiyuan
APPLICANT: Lodes, Michael
APPLICANT: Jones, Robert
APPLICANT: Benson, Darin
APPLICANT: Bailey, Patrick
APPLICANT: Barth, Brenda
APPLICANT: Tegliss, John
TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
FILE REFERENCE: 210121 514C1
CURRENT APPLICATION NUMBER: PCT/US02/32727
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 30992
SEQ ID NO 5526
LENGTH: 52
TYPE: PRT
ORGANISM: Propionibacterium
PCT-US02-32727-5526


```

: APPLICANT: Carter, Darriek
: APPLICANT: Barth, Brenda
: APPLICANT: Douglass, John
: TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
: FILE REFERENCE: 210121.514C1
: CURRENT APPLICATION NUMBER: US/09/978,825
: CURRENT FILING DATE: 2003-01-29
: NUMBER OF SEQ ID NOS: 30992
: SEQ ID NO 54926
: LENGTH: 52
: TYPE: PRT
: ORGANISM: Propioni acnes
: US-09-978-825-5526

Query Match
Best Local Similarity 66.7%, Score 4, DB 5, Length 52,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 2 VRWH 5
Db 26 VRWH 29

RESULT 7
US-10-057-498-5526
: Sequence 5526, Application US/10057498
: GENERAL INFORMATION:
: APPLICANT: Mitcham, Jennifer
: APPLICANT: Skeiky, Yasir
: APPLICANT: Persing, David
: TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
: FILE REFERENCE: 210121.514
: CURRENT APPLICATION NUMBER: us/10/057,498
: CURRENT FILING DATE: 2003-04-20
: NUMBER OF SEQ ID NOS: 29212
: SEQ ID NO 5526
: LENGTH: 52
: TYPE: PRT
: ORGANISM: Propioni acnes
: US-10-057-498-5526

Query Match
Best Local Similarity 66.7%, Score 4, DB 6, Length 52,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 2 VRWH 5
Db 26 VRWH 29

RESULT 8
PCT-US02-32727-7189
: Sequence 7189, Application PC/TUS0232727
: GENERAL INFORMATION:
: APPLICANT: Mitcham, Jennifer
: APPLICANT: Skeiky, Yasir
: APPLICANT: Persing, David
: APPLICANT: Phatla, Ajay
: APPLICANT: Maisonneuve, Jean Francois
: APPLICANT: Zhang, Yanni
: APPLICANT: Wang, Siqing
: APPLICANT: Jen, Shyian
: APPLICANT: Lodes, Michael
: APPLICANT: Benson, Darin
: APPLICANT: Jones, Robert
: APPLICANT: Carter, Darriek
: APPLICANT: Barth, Brenda
: APPLICANT: Douglass, John
: TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
: FILE REFERENCE: 210121.514C1
: CURRENT APPLICATION NUMBER: PCT/US02/32727
: CURRENT FILING DATE: 2002-10-11
: NUMBER OF SEQ ID NOS: 30992
```

```

: SEQ ID NO 7189
: LENGTH: 56
: TYPE: PRT
: ORGANISM: Propioni acnes
: PCT-US02-32727-7189

Query Match
Best Local Similarity 66.7%, Score 4, DB 3, Length 56,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 2 VRWH 5
Db 20 VRWH 23

RESULT 9
US-09-978-825-7189
: Sequence 7189, Application US/09978825
: GENERAL INFORMATION:
: APPLICANT: Mitcham, Jennifer
: APPLICANT: Skeiky, Yasir
: APPLICANT: Persing, David
: APPLICANT: Bhatla, Ajay
: APPLICANT: Maisonneuve, Jean Francois
: APPLICANT: Zhang, Yanni
: APPLICANT: Wang, Siqing
: APPLICANT: Jen, Shyian
: APPLICANT: Lodes, Michael
: APPLICANT: Benson, Darin
: APPLICANT: Jones, Robert
: APPLICANT: Carter, Darriek
: APPLICANT: Barth, Brenda
: APPLICANT: Douglass, John
: TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
: FILE REFERENCE: 210121.514C1
: CURRENT APPLICATION NUMBER: US/09/978,825
: CURRENT FILING DATE: 2003-01-29
: NUMBER OF SEQ ID NOS: 30992
: SEQ ID NO 7189
: LENGTH: 56
: TYPE: PRT
: ORGANISM: Propioni acnes
: US-09-978-825-7189

Query Match
Best Local Similarity 66.7%, Score 4, DB 5, Length 56,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 2 VRWH 5
Db 20 VRWH 23

RESULT 10
US-10-057-498-7189
: Sequence 7189, Application US/10057498
: GENERAL INFORMATION:
: APPLICANT: Mitcham, Jennifer
: APPLICANT: Skeiky, Yasir
: APPLICANT: Persing, David
: APPLICANT: Phatla, Ajay
: APPLICANT: Maisonneuve, Jean Francois
: APPLICANT: Zhang, Yanni
: APPLICANT: Wang, Siqing
: APPLICANT: Jen, Shyian
: APPLICANT: Lodes, Michael
: APPLICANT: Benson, Darin
: APPLICANT: Jones, Robert
: APPLICANT: Carter, Darriek
: APPLICANT: Barth, Brenda
: APPLICANT: Douglass, John
: TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V
: FILE REFERENCE: 210121.514
: CURRENT APPLICATION NUMBER: US/10/057,498
: CURRENT FILING DATE: 2003-04-20
: NUMBER OF SEQ ID NOS: 30992
: SEQ ID NO 7189
: LENGTH: 56
: TYPE: PRT
: ORGANISM: Propioni acnes
: US-10-057-498-7189

Query Match
Best Local Similarity 66.7%, Score 4, DB 6, Length 56,
Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0;
```

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 VRWH 5
DB 11 VRWH 14

RESULT 11
PCT-US02 32727-11626

Sequence 11626, Application FC/TUS0232727
GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
APPLICANT: Bhatia, Ajay
APPLICANT: Maisonneuve, Jean Francois
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Siying
APPLICANT: Jen, Chyian
APPLICANT: Lodes, Michael
APPLICANT: Benson, Darin
APPLICANT: Carter, Derrick
APPLICANT: Barth, Brenda
APPLICANT: Douglas, John
TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
FILE REFERENCE: 210121.51401
CURRENT APPLICATION NUMBER: PCT/US02/32727
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 30992
SEQ ID NO 11626
LENGTH: 58
TYPE: PRT
ORGANISM: Propionibacterium
PCT US02 32727 11626

Query Match 66.7%; Score 4; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 20+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWH 5
DB 11 VRWH 14

RESULT 12
US 09-978-825-11626

Sequence 11626, Application US/09978825
GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
APPLICANT: Bhatia, Ajay
APPLICANT: Maisonneuve, Jean Francois
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Siying
APPLICANT: Jen, Chyian
APPLICANT: Lodes, Michael
APPLICANT: Benson, Darin
APPLICANT: Carter, Derrick
APPLICANT: Barth, Brenda
APPLICANT: Douglas, John
TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
FILE REFERENCE: 210121.51401
CURRENT APPLICATION NUMBER: US/09/978,825
CURRENT FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 30990
SEQ ID NO 11626
LENGTH: 58
TYPE: PRT
ORGANISM: Propionibacterium
US 09-978-825-11626

Query Match 66.7%; Score 4; DB 5; Length 58;
Best Local Similarity 100.0%; Pred. No. 20+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWH 5
DB 11 VRWH 14

RESULT 13
US-10-057-498-11626

Sequence 11626, Application US/10057498
GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
APPLICANT: Bhatia, Ajay
APPLICANT: Maisonneuve, Jean Francois
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Siying
APPLICANT: Jen, Chyian
APPLICANT: Lodes, Michael
APPLICANT: Benson, Darin
APPLICANT: Carter, Derrick
APPLICANT: Barth, Brenda
APPLICANT: Douglas, John
TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
FILE REFERENCE: 210121.514
CURRENT APPLICATION NUMBER: US/10/057,498
CURRENT FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 22212
SEQ ID NO 11626
LENGTH: 58
TYPE: PRT
ORGANISM: Propionibacterium
US 10-057-498-11626

Query Match 66.7%; Score 4; DB 5; Length 58;
Best Local Similarity 100.0%; Pred. No. 20+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWH 5
DB 11 VRWH 14

RESULT 14
PCT US02 32727 24772

Sequence 24772, Application FC/TUS0232727
GENERAL INFORMATION:

APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
APPLICANT: Bhatia, Ajay
APPLICANT: Maisonneuve, Jean Francois
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Siying
APPLICANT: Jen, Chyian
APPLICANT: Lodes, Michael
APPLICANT: Benson, Darin
APPLICANT: Carter, Derrick
APPLICANT: Barth, Brenda
APPLICANT: Douglas, John
TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
FILE REFERENCE: 210121.51401
CURRENT APPLICATION NUMBER: PCT/US02/32727
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 30992
SEQ ID NO 24772
LENGTH: 64
TYPE: PRT
ORGANISM: Propionibacterium
PCT US02 32727 24772

Query Match 66.7%; Score 4; DB 1; Length 64;
Best Local Similarity 100.0%; Pred. No. 20+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWH 5
DB 11 VRWH 14

RESULT 15

US-09-978-825-24772

/ Sequence 24772, Application US/09979925

/ GENERAL INFORMATION:

/ APPLICANT: Mitcham, Jennifer

/ APPLICANT: Skeiky, Yasir

/ APPLICANT: Persing, David

/ APPLICANT: Bhatia, Ajay

/ APPLICANT: Maisonneuve, Jean Francois

/ APPLICANT: Zhang, Yanni

/ APPLICANT: Wang, Siqing

/ APPLICANT: Jen, Shyian

/ APPLICANT: Lodes, Michael

/ APPLICANT: Benson, Darin

/ APPLICANT: Jones, Robert

/ APPLICANT: Carter, Darrick

/ APPLICANT: Barth, Brenda

/ APPLICANT: Douglass, John

/ TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes V

/ FILE REFERENCE: 210121.514C1

/ CURRENT APPLICATION NUMBER: US/09/978,825

/ CURRENT FILING DATE: 2003-01-29

/ NUMBER OF SEQ ID NOS: 30992

/ SEQ ID NO 24772

/ LENGTH: 64

/ TYPE: PRT

/ ORGANISM: Propionl acnes

US-09-978-825-24772

Query Match

Best Local Similarity 66.7%; Score 4; DB 5; Length 64;

Matches 4; Conservatave 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VRWH 5

Db 2 VRWH 5

Search Completed: February 20, 2003, 11:42:48
Job time : 17.6667 secs

GenCore version 5.1.3
Copyright (c) 1992 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:24:02, Search time 12.69 sec
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-2
Perfect score: 6
Sequence: 1 WVPWHF 6

Scoring table: OLIGOEX
Gapop 60.0, Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 48 summaries

Database: PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred No is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the local score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	5	83.3	301	G83352	transcription regu
2	5	83.3	370	A96741	hypothetical prote
3	5	83.3	799	T48889	serine/threonine p
4	4	66.7	52	A95009	hypothetical prote
5	2	66.7	52	E97880	hypothetical prote
6	4	66.7	62	JC6155	viral capsid prote
7	2	66.7	95	S10796	conserved hypothet
8	4	66.7	98	AG3064	hypothetical prote
9	4	66.7	102	A99800	hypothetical prote
10	4	66.7	103	T27486	hypothetical prote
11	4	66.7	106	T44499	hypothetical prote
12	4	66.7	119	MMXP19	sigma tms protein
13	4	66.7	119	P34829	sigma tms protein
14	4	66.7	122	T04366	probable peroxidase
15	4	66.7	133	P97400	hypothetical prote
16	4	66.7	133	AC2619	conserved hypothet
17	4	66.7	135	D81704	hypothetical prote
18	4	66.7	149	GGIME	globin - river lam
19	4	66.7	149	GGIMS	globin V (validae
20	4	66.7	149	A26042	globin I - sea lam
21	4	66.7	149	B26042	globin III - sea l
22	4	66.7	149	S13458	hemoglobin - south
23	4	66.7	149	S13459	hemoglobin - south
24	4	66.7	149	S13460	hemoglobin - south
25	4	66.7	162	F82493	hypothetical prote
26	4	66.7	169	F75253	conserved hypothet
27	4	66.7	170	S64498	regulatory protein
28	4	66.7	171	AD2457	hypothetical prote
29	4	66.7	184	A71086	conserved hypothet

30	4	66.7	197	A92556	c-type cytochrome
31	4	66.7	206	T40138	hypothetical prote
32	4	66.7	208	H62350	conserved hypothet
33	4	66.7	209	C97377	HD superfamily hyn
34	4	66.7	210	D84016	hypothetical prote
35	4	66.7	215	T36900	probable transcrip
36	4	66.7	221	D97161	probable enzyme w
37	4	66.7	226	C96355	protein P101111
38	4	66.7	232	H75422	probable phosphatase
39	4	66.7	233	H71040	conserved hypothet
40	4	66.7	241	F72385	hypothetical prote
41	4	66.7	244	T15996	hypothetical prote
42	4	66.7	249	T12463	hypothetical prote
43	4	66.7	253	T03453	hypothetical prote
44	4	66.7	257	C98662	conserved hypothet
45	4	66.7	257	C83082	hypothetical prote

ALIGNMENTS

RESULT 1
G83352
transcription regulator MtlR PA2337 (imported) - Pseudomonas aeruginosa (strain PA01)
C/Species: Pseudomonas aeruginosa
C/Data: 15-Sep-2000 #seq:1000 #rev:1000 #ext_change:31-Dec-2000
C/Accession: G83352
P,Stover, C.F., Pham, X.Q., Erwin, M.L., Milt, P.H., S.E., Winkler, P., H. Loff, M. L. Br
adman, S.; Yuan, Y.; Brody, L.B.; Coulter, S.N.; Folger, K.P.; Kae, A.; Lathig, K.; Loff,
.; Loff, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathob
A/Reference number: A82950; MUP:00437337; PMID:10984043
A/Accession: G83352
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-301 <STO>
A/Cross-references: JB A80944; JP A81040; NT J73391; NM A84337; ZEP:000141
A/Experimental source: strain PA01
A/Genetics:
A/Gene: mtlR; PA2337

Query Match	83.3%	Score 5	DB 2	Length 301					
Best Local Similarity	100.0%	Pred. No. 17							
Matches	5	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	2	VRWHF 6							
Db	35	VRWHF 39							

RESULT 2
A96741
hypothetical protein F14003.02 (imported) - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Data: 02-Mar-2001 #seq:1000 #rev:1000 #ext_change:31-Mar-2001
C/Accession: A96741
P,Theologis, A.; Ecker, J.R.; Falm, C.C.; Federoff, N.A.; Fink, G.; Whitte, O.; Alonso,
Chin, J.W.; Chung, M.F.; Cline, D.; Cowley, K.P.; Givay, A.F.; Givay, J.W.; Givay, R.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 916-920, 2000
A,Authors: Hunter, J.L., Jenkins, G., Johnson-Hopson, C., Khan, S., Khaykin, E., Kim, C.,
C.A., Li, J.H., Li, Y., Lin, X., Liu, S.X., Liu, Z.A., Lures, D.S., Maiti, R., Marziani,
Pizarro, M., Pooney, T.; Powley, D.; Sakano, H.
A,Authors: Salzman, S.L., Schwartz, J.P., Shinn, P., Southwick, A.M., Sun, H., Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis
A/Reference number: A81411; MUP:01016713; PMID:11220712
A/Accession: A96741
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-370 <STO>
A/Cross-references: GB A809173; NT J73391; NM A84337; ZEP:000141

A/Cross-references: EMBL:Y00060, MID:340038, PIRN:CA37626.1, PIR:J1334248
R.Borges, A.J. Hawkins, C.F. Packman, L.O. Perham, R.N.
Eur. J. Biochem. 194, 95-102, 1990
A/Title: Cloning and sequence analysis of the genes encoding the dihydrodipicoyl acetyl
Bacillus stearothermophilus.
A/Reference number: S14838; MID:41071217; PMID:2252629
A/Accession: T46887
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-95 <BOR>
A/Cross-references: EMBL:X50663, MID:340038; PIRN:CA37626.1; PIR:J1334248
A/Experimental source: strain NCAL503
A/Superfamily: Methanobacterium thermophilum conserved hypothetical protein MTH10

Query Match 66.7%; Score 4; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VRWH 5
|||||
Db 5 VRWH 8

RESULT 8
AG3064
hypothetical protein Atub13 (fragment) - Agrobacterium tumefaciens (strain ATCC 27081)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C/Accession: AG3064
R/Moore, P.W.; Gubral, J.C.; Kaul, P.; Marks, D.; Cheng, L.; Ward, G.E.; Chen, Y.; Wroblewski, G.; Gillet, W.; Grant, C.; Gonthier, D.; Kutyavin, T.; Levy, P.; Li, M.; McClellan, J.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoon, H.; Tay, Y.; Riddle, P.; Jung, M.; Krespan, W.; Parry, M.; Gordon Kamm, Steve, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
A/Reference number: AB2577; PMID:11743193
A/Accession: AG3064
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-98 <KUP>
A/Cross-references: GB:AF098699; PIRN:AA14493.1, PIR:J17742997, GSPDB:G000187
A/Experimental source: strain C58 (Dupont)
A/Genetics:
A/Gene: Atub413
A/Map position: Linear Chromosome

Query Match 66.7%; Score 4; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VRWH 5
|||||
Db 3 VRWH 6

RESULT 9
A99800
hypothetical protein EGS1369 (fragment) - Escherichia coli (strain O157:H7, substrain R1)
C/Species: Escherichia coli
C/Date: 19-Jul-2001 #sequence_revision 19-Jul-2001 #text_change 19-Jul-2001
C/Accession: A99800
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Furukawa, F.; Ishii, F.; Yokoyama, F.; Han, C.G.; Gotohara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A/Reference number: A99800; MID:21156231; PMID:11258796
A/Accession: A99800
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-102 <HAV>
A/Cross-references: GB:BA000007; PIRN:PA34792.1; PIR:J13360829, GSPDB:G000154
A/Experimental source: strain O157:H7, substrain R1MD 0509952

A/Genetics:
A/Gene: EGS1369

Query Match 66.7%; Score 4; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RWHF 6
|||||
Db 39 RWHF 42

RESULT 10
T27486
hypothetical protein Y9C2UA.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 10-Mar-1993 #sequence_revision 10-Mar-1993 #text_change 04-Mar-2000
C/Accession: T27486
R/Smye, R.
submitted to the EMBL Data Library, December 1998
A/Reference number: Z20374
A/Accession: T27486
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-103 <WIL>
A/Cross-references: EMBL:AL094941, PIRN:CA10136.1, PIR:J1334248
A/Experimental source: clone Y9C2UA
A/Genetics:
A/Gene: CESP:Y9C2UA.2
A/Map position 2
A/Introns: 21/2
A/Superfamily: Caenorhabditis elegans hypothetical protein Y9C2UA.2

Query Match 66.7%; Score 4; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RWHF 6
|||||
Db 53 RWHF 56

RESULT 11
T44499
hypothetical protein (fragment) - Thermomonospora fusca (fragment)
C/Species: Thermomonospora fusca
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C/Accession: T44499
R/Irwin, D.T.; Zhang, S.; Wilson, D.P.
submitted to the EMBL Data Library, April 1999
A/Description: Characterization of a Thermomonospora fusca family 42 cellulase E6.
A/Reference number: Z22783
A/Accession: T44499
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-106 <IRW>
A/Cross-references: EMBL:AF144563; PIRN:AAD39950.1
A/Experimental source: strain YX

Query Match 66.7%; Score 4; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WVRW 4
|||||
Db 14 WVRW 17

RESULT 12
MNXR1B
sigma IBS protein - reovirus type 1 (strain Lang)
A/Alternative names: nonstructural protein
C/Species: reovirus type 1

A/Note: host Homo sapiens (man)
 C/Date: 30-Jun-1998 #sequence_revision 30-Jun-1998 #text_change 16-Jul-1999
 C/Accession: B26436
 R/Munemitsu, S.M.; Atwater, J.A.; Samuel, C.E.
 Biochem. Biophys. Res. Commun. 110, 528-534, 1993
 A/Title: Biochemical and molecular characterization of the sigma 1bns protein and the sigma 1bns peptide sigma 1bns.
 A/Reference number: A90129; PMID:9794999; PMID:2420558
 A/Accession: B26436
 A/Molecule type: mRNA
 A/Residues: 1-119 <MCN>
 A/Cross-references: GB:M470; NID:933344; PID:AA41274; PID:9333746
 A/Note: the authors translated the codon CAA for residue 5 as Gln
 C/Genetics:
 A/Map position: segment S1
 C/Superfamily: reovirus sigma 1bns protein
 C/Keyword: nonstructural protein

Query Match 66.7% Score 4; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. NO. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRWM 4
 ||||
 DB 50 VRWM 53

RESULT 13

B26436

sigma 1bns protein - reovirus type 1 (strain Lang)

N/Alternate names: nonstructural protein

C/Species: reovirus type 1

A/Note: host Homo sapiens (man)

C/Date: 30-Mar-1991 #sequence_revision 18-Mar-1991 #text_change 21-Sep-1993

C/Accession: B34829

R/Milner, M.L.; Demich, T.S.; Fields, B.N.

J. Virol. 64, 2976-2989, 1990

A/Title: Characterization of the reovirus sigma 1bns protein and the sigma 1bns peptide

A/Reference number: A34829; PMID:9024445; PMID:2335823

A/Accession: B34829

A/Status: preliminary

A/Molecule type: genomic RNA

A/Residues: 1-119 <MIF>

A/Cross-references: GB:M3363; NID:933342; PID:AA41243.1; PID:9333654

C/Genetics:

A/Map position: segment S1

C/Superfamily: reovirus sigma 1bns protein

Query Match 66.7% Score 4; DB 2; Length 119;
 Best Local Similarity 100.0%; Pred. NO. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRWM 4
 ||||
 DB 50 VRWM 53

RESULT 14

T04366

Probable peroxidase (EC 1.11.1.7), cationic - rice (fragment)

C/Species: Oryza sativa (rice)

C/Date: 23-Apr-1993 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999

C/Accession: T04366

R/Kim, C.Y.; Jeon, S.Y.; Choe, M.S.; Cheong, Y.H.; Han, C.D.; Cho, M.J.

submitted to the EMBL Data Library, March 1997

A/Description: Isolation and characterization of cDNAs inducible rice genes by a fungal

A/Reference number: Z15314

A/Accession: T04366

A/Status: preliminary; translated from cDNA/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-122 <KIM>

A/Cross-references: EMBL:U95217; NID:92443456; PID:AA57138.1; PID:92443459

C/Genetics:

A/Genes: PRX1
 C/Superfamily: peroxidase
 C/Keywords: oxidoreductase

Query Match 66.7% Score 4; DB 2; Length 122;
 Best Local Similarity 100.0%; Pred. NO. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWM 5
 ||||
 DB 50 VRWM 53

RESULT 15

B97400

Hydrolytic protein A97400 (unpublished) Agrobacterium tumefaciens strain 354, 1990

C/Species: Agrobacterium tumefaciens

C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C/Accession: B97400

R/Jalil, B.; Hukie, S.; Sattar, S.; Miller, N.; Blackland, N.; Qurollo, B.; Soliman, A.; Liu, F.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markovic, B.

Science 294, 2323-2328, 2001

A/Title: Genome sequence of the plant pathogen and biotechnological agent Agrobacterium tumefaciens

A/Reference number: A97359; PMID:11743194

A/Accession: B97400

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-133 <KUR>

A/Cross-references: GB:M1744; EMBL:U95217; NID:92443456; PMID:9024445

C/Genetics:

A/Genes: AGR_C_590

A/Map position: circular chromosome

Query Match 66.7% Score 4; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. NO. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWM 5
 ||||
 DB 107 VRWM 110

Search completed: February 20, 2003, 11:09:39
 Job time: 13 secs

GenCore version 5.1.3
Copyright (c) 1993 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:20:17, Search time 11.333 seconds
(without alignments)
21,058 Million cell updates/sec

Title: US-09-912-414-2

Perfect score: 6

Sequence: 1 WVRMHF 6

Scoring table: OLIGOX

Gapop 60.0, Gdext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: SwissProt_40*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	5	83.3	372	GP45_HUMAN	Q95Y93 homo sapien
2	5	83.3	373	GP45_MOUSE	Q9eqg4 mus musculu
3	5	83.3	799	AFSK_STRGO	P54741 streptomyce
4	5	83.3	807	AFSK_STRGR	P54742 streptomyce
5	4	66.7	95	YPPA_BACST	P21878 bacillus st
6	4	66.7	119	VSIS_PECVL	P07938 reovirus (t
7	4	66.7	149	GLR1_MOPMP	P21197 mordacia mo
8	4	66.7	149	GLR1_PETMA	P09667 petromyzon
9	4	66.7	149	GLR2_MOPMP	P21198 mordacia mo
10	4	66.7	149	GLR2_PETMA	P21193 petromyzon
11	4	66.7	149	GLB3_MOPMP	P21199 mordacia mo
12	4	66.7	149	GLB3_PETMA	P09968 petromyzon
13	4	66.7	149	GLB5_PETMA	P22208 petromyzon
14	4	66.7	149	GLR_TAMET	P22207 lametria fi
15	4	66.7	170	CRP4_YEAST	P37267 saccharomyc
16	4	66.7	195	TNFT_MOUSE	Q55237 mus musculu
17	4	66.7	197	DSBE_XYLEA	Q9pan4 xylella fas
18	4	66.7	208	YQED_BACSU	P54449 bacillus su
19	4	66.7	215	UP2P_POVIN	P49220 hos taurus
20	4	66.7	223	PQCC_HUMAN	O15173 homo sapien
21	4	66.7	244	YVGA_CABER	P22067 caenorhabdi
22	4	66.7	286	PURU_CORSP	Q46339 corynebacte
23	4	66.7	295	MOPA_PSPPU	Q02198 pseudomonas
24	4	66.7	298	YIHV_ECOLI	P32143 escherichia
25	4	66.7	310	DNUL_MYCGE	P47248 mycoplasma
26	4	66.7	322	ASTE_ECOLI	P76215 escherichia
27	4	66.7	325	UL76_HCVVA	P16725 human cytom
28	4	66.7	338	PDXA_PALSO	P58714 ralstonia s
29	4	66.7	399	YVET_MYCTT	Q54697 xylella fast
30	4	66.7	403	YVHM_PHSIN	P55480 rhizobium s
31	4	66.7	418	PHAA_BACHP	Q94019 bacillus ha
32	4	66.7	422	TERI_PSPPF	P54308 bacteriopho
33	4	66.7	436	TER4_MOUSE	Q91236 mus musculu

34	4	66.7	440	1	SOCA_HUMAN	Q46X5 homo sapien
35	4	66.7	450	1	YOE9_PSEAE	Q9HY93 pseudomonas
36	4	66.7	462	1	SYTM_YEAST	P22316 saccharomy
37	4	66.7	472	1	PBB_ESCPE	P21948 escherichia
38	4	66.7	475	1	ETS6_PROME	P24776 drosophila
39	4	66.7	486	1	YDF1_ECOLI	P77260 escherichia
40	4	66.7	507	1	TRA6_PSEAE	Q57541 pseudomonas
41	4	66.7	508	1	TRAT_CHEHE	Q40092 cheila elact
42	4	66.7	511	1	FAST_MOUSE	Q91189 mus musculu
43	4	66.7	545	1	FAST_HUMAN	Q14236 homo sapien
44	4	66.7	583	1	STS_HUMAN	P08842 homo sapien
45	4	66.7	611	1	SNF1_CANCA	Q00372 candida gla

ALIGNMENTS

RESULT 1

GP45_HUMAN

ID GP45_HUMAN STANDARD; PRT; 372 AA.

AC Q9Y5Y3;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable G protein-coupled receptor GPR45 (PSP24-alpha) (PSP24-1).

GN GPR45.

OS Homo sapiens (Human).

OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,

OC Mammalia, Eutheria, Primates, Catarrhini, Homidae, Homo.

OX NCBI_TaxID=9606;

RN [1]

SEQUENCE FROM N.A.

FX MEDLINE:2015652, PubMed 10036181,

EA Marchese A., Sawadsky M., Nguyen T., Cheng F., Heng H.H., Nowak T.,

RA Im D.-S., Lynch K.P., George S.P., O'Dowd B.F.;

RT "Discovery of three novel orphan G-protein-coupled receptors.";

RL Genomics 56:12-21(1999).

CC -!- FUNCTION: Orphan receptor. May play a role in brain function.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- TISSUE SPECIFICITY: Expressed in brain; detected in the basal

CC forebrain, frontal cortex, and caudate, but not in thalamus,

CC hippocampus, or putamen.

CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.ebi.ac.uk/submit/>

CC or send an email to license@ebi.ac.uk).

CC EMPI: A119266, A1P2356.1; -.

DR GDB: 604838; -.

DR GDB: 604838; -.

DR InterPro: IPR002276; GPCR_Fhodpsn.

DR Pfam: PF00001; 7tm_1; 1.

DR PRINTS: PR00237; GPCR_Fhodpsn.

DR PROSITE: PS00237, G-PROTEIN_RECEP_F1_1, FALSE_NEG.

DE PROSITE: PS00237, G-PROTEIN_RECEP_F1_2, 1.

DE G protein-coupled receptor, transmembrane, G-protein.

FT DOMAIN 1 38

FT TRANSMEM 39 59

FT DOMAIN 60 75

FT TRANSMEM 76 96

FT DOMAIN 110 130

FT TRANSMEM 131 149

FT DOMAIN 150 172

FT TRANSMEM 171 198

FT DOMAIN 199 219

FT TRANSMEM 220 268

FT DOMAIN 269 289

FT TRANSMEM 289 372


```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcements/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D45392; EAA09229.2;
DR FMBL: A1392178; CAC09412.1;
DR FMBL: A1392178; CAC09412.1;
DR PhosSite; P54741;
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR01245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00223; S_TKc; 1.
DR SMART; SM00219; TYRKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR TRANSFERASE; Serine/threonine-protein kinase; ATP binding;
KW Phosphorylation; complete; prot.com
FT DOMAIN 16 271 PROTEIN KINASE.
FT NF_BIND 22 30 ATP (BY SIMILARITY).
FT BINDING 44 44 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 794 AA; 4374 MW; 4E5EDED41B9F5E5E CDS64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 794;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWHF 6
Db 628 VRWHF 632

RESULT 4
AFSK_STRGR
ID AFSK_STRGP STANDARD; PRT; 907 AA.
AC p64742;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein kinase afsk (Ec 2.7.1.1 -)
GN AFSK.
OC Streptomyces gilvusc.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycinales; Streptomycesaceae; Streptomyces
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE 261878; PubMed=2203674.
RA Ueda K., Uemayama T., Beppu T., Horinouchi S.,
RT "The aerial mycelium defective phenotype of Streptomyces gilvusc
RT resulting from A-factor deficiency is suppressed by a Ser/Thr kinase
RT of S. coelicolor A3(2).";
RL Gene 169:91-95(1996).
CC -1- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
CC PHOSPHORYLATING, ON BOTH SEP AND THP, THE AFSP GLOBAL REGULATORY
CC PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM (BY
CC SIMILARITY).
CC -1- PFM. AUTOPHOSPHORYLATED AT THP AND SEP RESIDUES (BY SIMILARITY)
CC -1- SIMILARITY. PFMONS TO THE SEP/THP FAMILY OF PROTEIN KINASES
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D45246; BAA08203.1;
DR PhosSite; P54742;
```

```
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 16 272 PROTEIN KINASE.
FT NF_BIND 22 30 ATP (BY SIMILARITY).
FT BINDING 44 44 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 907 AA; 46731 MW; 4E5EDED41B9F5E5E CDS64;

Query Match
Best Local Similarity 83.3%; Score 5; DB 1; Length 807;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWHF 6
Db 635 VRWHF 639

RESULT 5
YFDA_BACST
ID YFDA_FANCT STANDARD; PRT; 95 AA.
AC F21878;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in yfha 5' region (ORF1) (Fragment).
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCA 1503,
RX MEDLINE=26345939, PubMed=2203674,
RA Hawkins C.F., Borges A., Perham R.N.;
RT "Cloning and sequence analysis of the genes encoding the alpha and
RT beta subunits of the E1 component of the pyruvate dehydrogenase
RT multienzyme complex of Bacillus stearothermophilus.";
RL Burt J. Biochem. 191:337-346(1990).
CC -1- SIMILARITY. HEAVENS TO THE DEF/YFHA/YFHA/YFHA/ YFHA/ YFHA/ YFHA
CC (B.SUBTILIS) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53560; CAA37626.1;
DR PIR; S10796; S10796.
DR InterPro; IPR001454; Hlyase/hydrolase.
DR InterPro; IPR000150; Hypothet_cof.
DR Pfam; PF00702; Hydrolase; 1.
DR PROSITE; PS01228; GCF_1; PAFTHL.
DR PROSITE; PS01229; GCF_2; 1.
KW Hypothetical protein.
FT NON_TPR 1
SQ SEQUENCE 95 AA; 10441 MW; 811E9487321A74E3 CDS64;

Query Match
Best Local Similarity 66.7%; Score 4; DB 1; Length 95;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWHF 5
Db 5 VRWHF 8
```

```

RESULT 6
VOID_RECORD
ID VOID_RECORD STANDARD FRT: 119 AA.
AC P07038;
DT 01-AUG-1998 (Rel. 08, Created)
DI 01-AUG-1998 (Rel. 08, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Sigma I-2 protein.
OS Picovirus (Type 1 / strain Lang).
CC Viruses; dsRNA viruses; Reoviridae; Orthoreovirinae.
CX NCBI TaxId=10884;
RN 113;
RP SEQUENCE FROM N.A.
PX MEDLINE:87048783, PubMed:1435569;
RA Muenchitsu S.M., Alwaleel J.A., Samuel G.B.;
RT "Biosynthesis of reovirus-specified polypeptides Molecular CDNA cloning and nucleotide sequence of the coding strand."
RT "Genetic characterization of RNA which encodes the minor capsid polypeptide sigma Ia from the reovirus strain Lang."
RL Biochem Biophys Res Commun 140:504-514(1986);
RN 121;
RP SEQUENCE FROM N.A.
RX MEDLINE:90214415, PubMed:2335933;
RA Nibert M.L., Dermody T.S., Fields B.M.;
RT "Structure of the reovirus cell-attachment protein: A model for the domain organization of sigma 1."
RL Virol. 64:2976-2989(1990);
RN 131;
RP SEQUENCE FROM N.A.
RX MEDLINE:90211159, PubMed:1993945;
RA Abdel-Malek Z.W., Chapiro P.A., Wiener J.R., Toklik W.F.;
RT "Sequence of the C1 gene of the three serotypes of reovirus."
RL Proc Natl Acad Sci U.S.A. 82:24-28(1985);
RN 132;
RP This article has been published by the National Institutes of Health between the Swiss Institute of Bioinformatics and the EMBL database. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities relating to this record is prohibited. See http://www.ebi.ac.uk/seqdb/doc/embldb.html
RN 133;
DR EMBL: M44773; AAA47243.1;
DR EMBL: M55963; AAA47243.1;
DR EMBL: M10260; AAA66878.1;
DR EMBL: E04459; NM0042;
DR InterPro: IPR003478; Sigma_1s;
DR Pfam: PF02454; Sigma_1s;
FT CONFIDENTIAL
SC SEQUENCE 119 AA; 1199 MW; 15555555555555555555 CRC64;

Query Match: 68.3%, Score 4, 18.1, Length 119, Best Local Similarity: 100.0%, 1199 MW, 11 Matches, 0 Misses, 0 Indels, 0 Gaps, Db 50 WORDS

RESULT 7
GIBI_MORMR
ID GIBI_MORMR STANDARD; PRT: 149 AA.
AC P21197;
DT 01-MAY-1991 (Rel. 18, Created)
DI 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-1988 (Rel. 41, Last annotation update)
DE Globin T.
OS Moraxella mordax (Southern hemisphere lamprey).
CC Fishery; Mollusca; Chordata; Craniata; Vertebrata; Hyperoartia; Retrayzomiflorae; Retrayzomiflorae; Moraxella

```

```

OK NCBI_Taxid:7755;
RN (1)
RP SEQUENCE.
RX MEDLINE 21248417, PubMed:2039605;
RA Haddad J., Vidal V., Rodewald K., Braunitzer G., Neuzil E.;
RT "The primary structure of the hemoglobins of a southern hemisphere
  lamprey (Mordacia mordax, Cyprinostomata)";
AC Biol. Chem. Hoppe-Seyler 377:49-56(1991).
CC 1 SUBUNIT: Monomer.
CC -1 SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR: S13458; S13458.
DR HEST: F02208; 2LH8.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin; 1.
DR PROSITE: PS01033; GLOBIN; 1.
KW Hemoglobin transport; Transport.
KW Sequence: 149 AA; 16338 MW; 680239531247F146 CRO64;

2 VPMH 5
|||
DB 70 VRWH 72

RESULT 8
GLOB1_PETMA STANDARD; PRT; 149 AA.
ID GLOB1_PETMA
AC P09967;
DE 01-MAR-1988 (Rel. 10, Created)
DE 01-MAR-1988 (Rel. 10, Last sequence update)
DE 01-DEC-1991 (Rel. 10, Last annotation update)
DE Globin 1.
OS Petromyzon marinus (Sea lamprey).
PE Eukaryota, Metazoa, Chordata, Vertebrata, Hyperoartia;
NC Petromyzonidae (Petromyzonidae); Petromyzon.
KW NCBI_Taxid:7757,
RN (1)
RP SEQUENCE.
RX MEDLINE 6724703, PubMed:3766015;
RA Haddad J., Rodewald K., Vidal V., Neuzil E., Braunitzer G.;
RT "Primary structure of the minor hemoglobins from the sea lamprey
  (Petromyzon marinus, Cyprinostomata)";
AC Biol. Chem. Hoppe-Seyler 368:145-154(1987).
CC 1 SUBUNIT: Monomer.
CC -1 MISCELLANEOUS: THIS IS ONE OF THE MINOR GLOBIN COMPONENT OF SEA
  LAMPREY.
CC -1 SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR: A10041; A6042.
DR HEST: F02208; 2LH8.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; Globin; 1.
DR PROSITE: PS01033; GLOBIN; 1.
KW Hemoglobin transport; Transport.
KW Sequence: 149 AA; 16338 MW; 680239531247F146 CRO64;

Query Match 5678; Score 4; DB 1; Length 149;
Result: 100.00; Freq. No. 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 VPMH 5
|||
DB 70 VRWH 73

RESULT 9
GLOB2_MORMR STANDARD; PRT; 149 AA.
ID GLOB2_MORMR
AC P21109;
DE 01-MAY-1991 (Rel. 18, Created)
DE 01-MAY-1991 (Rel. 18, Last sequence update)

```


DT 15-JUN-2002 (Rel 41, last annotation update)
DE Globin II.
OS Mordacia mordax (Southern hemisphere lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Mordacia.
OX NCBI_TaxID=7755;
RN [1]
RP SEQUENCE.
RX MEDLINE:9124417; PubMed:2029605;
RA Hombrades I., Vidal Y., Podewald K., Braunltzer G., Neuzil E.,
RT "The primary structure of the hemoglobins of a southern hemisphere
lamprey (Mordacia mordax, Cyclostomata).";
RL Biol. Chem. Hoppe-Seyler 372:49-56(1991).
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR: S13459; S13459.
DR HSSP: P02208; 2LHB.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin; 1.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 149 AA, 16410 MW, 16508 MW, 215602E04F8EE666 CP764;

Query Match 66.7%; Score 4; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VRWH 5
Db 70 VRWH 73

RESULT 10
GLB3 PETMA
ID GLB3 PETMA STANDARD; PRT; 149 AA.
AC Q91913;
DT 16-OCT-2001 (Rel 40, Created)
DT 16-OCT-2001 (Rel 40, last sequence update)
DT 16-OCT-2001 (Rel 40, last annotation update)
DE Globin II (Hemoglobin PM1').
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:20250908; PubMed=10789466;
RA Qiu Y., Maillet D.H., Knapp J., Olsen J.S., Riggs A.F.;
RT "Lamprey hemoglobin: Structural basis of the Bohr effect";
RL J. Biol. Chem. 275:13517-13528(2000).
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC
DR EMBL: AF248645; AAF67186.1;
DR HSSP: P02208; 2LHB.
DR InterPro: IPR000971; Globin.
DR InterPro: IPR002335; Myoglobin.
DR Pfam: PF00042; globin; 1.
DR PRINTS: PR00613; MYOGLOBIN.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
FT INIT MET 0
SQ SEQUENCE 149 AA, 16411 MW, 16509 MW, 215602E04F8EE666 CP764;

Query Match 66.7%; Score 4; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 40;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 VRWH 5
Db 70 VRWH 73

RESULT 11
GLB3 MORMR
ID GLB3 MORMR STANDARD; PRT; 149 AA.
AC P21159;
DT 01-MAY-1991 (Rel 18, Created)
DT 01-MAY-1991 (Rel 18, last sequence update)
DT 15-JUN-2002 (Rel 41, last annotation update)
DE Globin III.
OS Mordacia mordax (Southern hemisphere lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Mordacia.
OX NCBI_TaxID=7755;
RN [1]
RP SEQUENCE.
RX MEDLINE:9124417; PubMed:2029605;
RA Hombrades I., Vidal Y., Podewald K., Braunltzer G., Neuzil E.,
RT "The primary structure of the hemoglobins of a southern hemisphere
lamprey (Mordacia mordax, Cyclostomata).";
RL Biol. Chem. Hoppe-Seyler 372:49-56(1991).
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR: S13460; S13460.
DR HSSP: P02208; 2LHB.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin; 1.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 149 AA, 16508 MW, 215602E04F8EE666 CP764;

Query Match 66.7%; Score 4; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VRWH 5
Db 70 VRWH 73

RESULT 12
GLB3 PETMA
ID GLB3 PETMA STANDARD; PRT; 149 AA.
AC P09968;
DT 01-MAP-1989 (Rel 10, Created)
DT 01-MAP-1989 (Rel 10, last sequence update)
DT 16-OCT-2001 (Rel 40, last annotation update)
DE Globin III.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE.

RP
RX MEDLINE:87184913; PubMed:3566915;
RA Hombrades I., Podewald K., Allard M., Neuzil E., Braunltzer G.,
RT "Primary structure of the minor haemoglobins from the sea lamprey
(Petromyzon marinus, Cyclostomata).";
RL Biol. Chem. Hoppe-Seyler 368:145-154(1987)
CC -1- SUBUNIT: Monomer.
CC -1- MISCELLANEOUS: THIS IS ONE OF THE MINOR GLOBIN COMPONENT OF SEA
CC LAMPREY.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR: B26043; B26042
DR HSSP: P02209; 2LHB.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin; 1.
DR PROSITE: PS01033; GLOBIN; 1.


```

ID CBP4_YEAST STANDARD; PRT; 170 AA.
AC P17267;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CBP4 protein, mitochondrial precursor.
GN CBP4 OR YGR174C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi, Ascomycota, Saccharomycotina, Saccharomycetes;
OC Saccharomycetales, Saccharomycetaceae, Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94342301; Pubmed=8063753;
RA Crivellone M.D.;
RT "Characterization of CBP4, a new gene essential for the expression of
  ubiquinol-cytochrome c reductase in Saccharomyces cerevisiae."
  J Biol Chem 269:21284-21292(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Hebling U., Hofmann B., Delius H.;
RT Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR THE ASSEMBLY AND/OR STABILITY OF
  TRICHLINOL-CYTOCHROME C REDUCTASE. IT HAS A DIRECT EFFECT ON THE
  EFFECT OF SUPPRESSION OF THE PLECKE PROTEIN, CORE 4, CORE 5 AND
  APOCYPOTROPOME R; IT MAY EITHER BE INVOLVED IN POST-TRANSLATIONAL
  MODIFICATION OF THE SUBUNITS OR IN THE ASSEMBLY OF THE ENZYME.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL ASSOCIATED WITH THE INNER
  MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10700; AAA61566.1; -
DR EMBL; Z72959; CAA97200.1; -
DR PIR; A53928; A53928.
DR SCD; S0003406; CBP4.
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 2 MITOCHONDRION (POTENTIAL).
FT CHAIN 2 170 CBP4 PROTEIN.
FT CONFLICT 65 65 S -> F (IN REF. 1).
SQ SEQUENCE 170 AA; 20019 MW, 398702FADFD0R366R QPC64;

```

```

Query Match          66.7%; Score 4; DB 1; Length 170;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 WVRW 4
   ||||
Db 29 WVRW 32

```

Search completed: February 20, 2003, 11:24:56
Job time : 12.3333 secs



GenCore version 5.1.3
Copyright (c) 1993 2003 CompuGen Ltd.

OM Filter: partial search, using sw model

Run on February 20, 2003, 11:33:17, Search time 47 seconds
(without alignments)
26.876 Million cell updates/sec

Title: US-09-912-414-2

Perfect score: 6

Sequence: 1 WVRWHP 6

Scoring table: Gapped

Gapop 60 0, Gapext 60 0

Searched: 671586 seqs, 286847115 residues

Word size: 6

Total number of hits satisfying chosen parameters: 671586

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Fast processing: listing first 45 summaries

Database:

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriag:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARYS

Result No	Score	Query Match length DB	ID	Description
1	5	83.3	301 2 052770	052770 pseudomonas
2	5	83.3	301 16 0911E1	0911E1 pseudomonas
3	5	83.3	308 6 08WNV6	08WNV6 capra hircu
4	5	83.3	366 16 09FC71	09FC71 streptomyce
5	5	83.3	372 10 02M3G2	02M3G2 arabidopsis
6	5	83.3	376 4 043898	043898 homo sapien
7	5	83.3	407 10 08RXQ4	08RXQ4 arabidopsis
8	4	66.7	52 16 097T74	097T74 streptococ
9	4	66.7	62 12 09A151	09A151 kaposi's sa
10	4	66.7	64 16 09RPN7	09RPN7 thermocact
11	4	66.7	76 10 09ACT2	09ACT2 cryza sativ
12	4	66.7	83 11 099LX1	099LX1 mus musculu
13	4	66.7	91 12 046412	046412 grapevine v
14	4	66.7	94 17 08UOY1	08UOY1 pyrococcus
15	4	66.7	98 16 08USF9	08USF9 agrobacteri
16	4	66.7	102 16 08XCK1	08XCK1 escherichia

17	4	66.7	103 5 09XXT9	09XXT9 camphorhadi
18	4	66.7	104 2 0918F3	0918F3 rhodospiril
19	4	66.7	106 2 09XCL1	09XCL1 thermococ
20	4	66.7	106 17 08U1G4	08U1G4 pyrococcus
21	4	66.7	108 11 09D688	09D688 mus musculu
22	4	66.7	110 2 0939C2	0939C2 thalassoscu
23	4	66.7	118 11 09CMT2	09CMT2 mus musculu
24	4	66.7	118 14 08U1T6	08U1T6 thalassoscu
25	4	66.7	122 10 024523	024523 cryza sativ
26	4	66.7	123 4 09YXN1	09YXN1 homo sapien
27	4	66.7	129 10 0864F5	0864F5 cryza sativ
28	4	66.7	131 10 09M6T3	09M6T3 abies alba
29	4	66.7	131 10 09M6T1	09M6T1 abies alba
30	4	66.7	132 10 09QAF8	09QAF8 arabidopsis
31	4	66.7	133 16 08U1F8	08U1F8 agrobacteri
32	4	66.7	135 16 08K2G4	08K2G4 silene alba
33	4	66.7	135 16 08Y1S1	08Y1S1 talassoscu
34	4	66.7	139 2 0462G9	0462G9 talassoscu
35	4	66.7	141 11 0861S5	0861S5 mus musculu
36	4	66.7	145 2 0864J7	0864J7 pseudomonas
37	4	66.7	150 5 0250J4	0250J4 halloctis la
38	4	66.7	150 5 0250J2	0250J2 halloctis la
39	4	66.7	150 5 0250J1	0250J1 halloctis la
40	4	66.7	150 5 0250J4	0250J4 halloctis la
41	4	66.7	151 4 08W3J4	08W3J4 homo sapien
42	4	66.7	151 11 092P47	092P47 mus musculu
43	4	66.7	154 5 0251S6	0251S6 halloctis la
44	4	66.7	155 16 0823P6	0823P6 thalassoscu
45	4	66.7	161 4 09Y5S5	09Y5S5 homo sapien

ALIGNMENTS

RESULT 1

052770 PRELIMINARY; PRT; 301 AA.

ID 052770 AC 052770; 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Activator protein.

GN MTLR.

OS Pseudomonas fluorescens.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID:294; OX (1)

PP SEQUENCE FROM N A

RC STRAIN=DSM 50106; MEDLINE=08332715, PubMed=966063;

FX Brunken P., Hils M., Altenbuchner J., Mattes R.; "The mannitol utilization genes of pseudomonas fluorescens are regulated by an activator cloning, nucleotide sequence and expression of the mtlr gene";

RT Gene 215:19-27(1998).

RL [2]

FE SEQUENCE FROM N A.

RC STRAIN=DSM 50106; Brunken P., Altenbuchner J.; Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RL SIMILARITY: BELONGS TO THE ARAC/XYLs FAMILY OF TRANSCRIPTIONAL REGULATORS.

CC EMBL; AF047527; AAC34292.1; -.

DR Interpro; IPR000005; HTHARAC.

DR Pfam; PF00165; HTH_ARAC; 2.

DR PRINTS; PP00032; HTHARAC.

DR SMART; SMC0342; HTH_ARAC; 1

DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.

DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.

KW DNA-binding; Transcription regulation.

SQ SEQUENCE 301 AA; 14776 MW; 08012FAFE524DD06 CRC64;

Query Match: At 14; Score 5; E8 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VWRWH 6
|||||
Db 27 VWRWH 31

RESULT 5
Q9M9G2 PRELIMINARY; PRT; 370 AA.

AC Q9M9G2; PRELIMINARY; PRT; 370 AA.
DT 01-OCT-2000 (TREMBlrel 15, Created)
DT 01-OCT-2000 (TREMBlrel 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
DE F14023.22 protein.
GN F14023.22.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core Eudicotyledons; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsi
OX NCBI_TaxID=3702;

RN [1]
PP SEQUENCE FROM N.A.
PC TISSUE=BRAIN;
RA An S.;
RL Submitted (MAP-1997) to the EMBL/GenBank/DBJ databases
EF EMBL: U92642, AAB23884.1, ..
DR InterPro; IPR000276; GPCR_Phodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRPHODPSN.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 370 AA; 4456 MW; 44053388083000 CRO64;

Query Match: At 14; Score 5; E8 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VWRWH 5
|||||
Db 151 VWRWH 155

RESULT 6
Q43898 PRELIMINARY; PRT; 376 AA.

AC Q43898;
DT 01-JUN-1998 (TREMBlrel 06, Created)
DT 01-JUN-1998 (TREMBlrel 06, Last sequence update)
DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
DE High-affinity tyrosylphosphatidic acid receptor homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euhorata; Primates; Anthropo; Hominoidea; Homo
OX NCBI_TaxID=9606;

RN [1]
PP SEQUENCE FROM N.A.
PC TISSUE=BRAIN;
RA An S.;
RL Submitted (MAP-1997) to the EMBL/GenBank/DBJ databases
EF EMBL: U92642, AAB23884.1, ..
DR InterPro; IPR000276; GPCR_Phodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRPHODPSN.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 370 AA; 4456 MW; 44053388083000 CRO64;

Query Match: At 14; Score 5; E8 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VWRWH 5
|||||
Db 151 VWRWH 155

RESULT 6
Q43898 PRELIMINARY; PRT; 376 AA.

AC Q43898;
DT 01-JUN-1998 (TREMBlrel 06, Created)
DT 01-JUN-1998 (TREMBlrel 06, Last sequence update)
DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
DE High-affinity tyrosylphosphatidic acid receptor homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euhorata; Primates; Anthropo; Hominoidea; Homo
OX NCBI_TaxID=9606;

Query Match: At 14; Score 5; E8 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VWRWH 6
|||||
Db 98 VWRWH 102

RESULT 7
Q8RXQ4 PRELIMINARY; PRT; 407 AA.

AC Q8RXQ4; PRELIMINARY; PRT; 407 AA.
DT 01-JUN-2002 (TREMBlrel 21, Created)
DT 01-JUN-2002 (TREMBlrel 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
DE Hypothetical 43 kDa protein.
GN A1G71840.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core Eudicotyledons; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsi
OX NCBI_TaxID=3702;

RN [1]
PP SEQUENCE FROM N.A.
PC TISSUE=BRAIN;
RA An S.;
RL Submitted (MAP-1997) to the EMBL/GenBank/DBJ databases
EF EMBL: U92642, AAB23884.1, ..
DR InterPro; IPR000276; GPCR_Phodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRPHODPSN.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 370 AA; 4456 MW; 44053388083000 CRO64;

Query Match: At 14; Score 5; E8 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VWRWH 5
|||||
Db 160 VWRWH 164

RESULT 8
Q97T74 PRELIMINARY; PRT; 52 AA.

AC Q97T74;
DT 01-OCT-2001 (TREMBlrel 18, Created)
DT 01-OCT-2001 (TREMBlrel 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel 19, Last annotation update)
DE Hypothetical protein SP0077.
GN SP0077.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;

RN [1]
PP SEQUENCE FROM N.A.
PC TISSUE=BRAIN;
RA An S.;
RL Submitted (MAP-1997) to the EMBL/GenBank/DBJ databases
EF EMBL: U92642, AAB23884.1, ..
DR InterPro; IPR000276; GPCR_Phodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRPHODPSN.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 370 AA; 4456 MW; 44053388083000 CRO64;

Query Match: At 14; Score 5; E8 16; Length 166;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VWRWH 5
|||||
Db 160 VWRWH 164

RESULT 8
Q97T74 PRELIMINARY; PRT; 52 AA.

AC Q97T74;
DT 01-OCT-2001 (TREMBlrel 18, Created)
DT 01-OCT-2001 (TREMBlrel 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel 19, Last annotation update)
DE Hypothetical protein SP0077.
GN SP0077.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;

RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA MEDLINE 119323; PMID 1193310;
RA Tetselin H., Nelson K.E., Palsson L.T., Elson J.A., Read T.D.,
RA Petersen S., Heideberg J., Leroy S.J., Hoff H., Heideberg J.,
RA Durkin A., Tatum M., Kallay T.E., Nelson M., Fedorova N.,
RA Grayam L.A., White O., Salzberg S.D., Lewis M.B., Kadane D.,
RA Holtzapple E., Abouli H., Wolf A.M., Ullrich T.E., Hamon C.L.,
RA M. J. J. J., Hoff H., Hoff H., Hoff H., Hoff H., Hoff H.,
RA Holt J.E., Loftus R.D., Yang F., Smith H.O., Venter J.C.,
RA Dougherty R.A., Hollnagel O.P., Fraser J.M.,
RA "Complete genome sequence of a wild-type isolate of Streptococcus
RT pneumoniae."
RI Science 297:404-409 (2002).
DR EMBL, AF007522, AF007522.1,
DR TIGR, ST0077, .
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 32 AA; 7051 MW; 19046FFD749F34C CRC64;

Query Match 66.7%; Score 4; DB 16; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 3 RWHF 6
DB 19 RWHF 22

RESULT 9
Q98153
ID Q98153 PRELIMINARY; FRT; 62 AA.
AC Q98153;
DT 01-FEB-1997 (TREMBL) 62, Created;
DT 01-FEB-1997 (TREMBL) 62, Last sequence update;
DT 01-NOV-1998 (TREMBL) 62, Last annotation update;
DE Unidentifiedorf.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; DNA viruses; no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Gammaherpesvirinae;
OX NCBI_TaxID:37226;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR;
RA Zhang W., Wang H., Herndler B., Garon D.,
RA "Restricted expression of Kaposi sarcoma-associated herpesvirus (human
RA herpesvirus 8) protein in human sarcoma."
RL Proc. Natl. Acad. Sci. USA 95:1441-1446 (1998).
DR EMBL, U66521, AAC55373.1, .
SQ SEQUENCE 62 AA; 7047 MW; D7CAB11C1AA2B2D CRC64;

Query Match 66.7%; Score 4; DB 12; Length 62;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 2 VPWF 5
DB 31 VPWF 34

RESULT 10
Q986N7
ID Q986N7 PRELIMINARY; FRT; 61 AA.
AC Q986N7;
DT 01-JUN-2002 (TREMBL) 61, Created;
DT 01-JUN-2002 (TREMBL) 61, Last sequence update;
DT 01-JUN-2002 (TREMBL) 61, Last annotation update;
DE Hypothetical protein TFE2763.
OS Thermotoga maritima.
OC Bacteria; Firmicutes; Firmicutes; Thermotogae; Thermotogaceae;
OC Thermotoga maritima; Thermotoga maritima; Thermotoga maritima;

OX NCBI_TaxID:119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEAT / JCM11007;
RA MEDLINE 119323; PMID 1193310;
RA Tetselin H., Nelson K.E., Palsson L.T., Elson J.A., Read T.D.,
RA Petersen S., Heideberg J., Leroy S.J., Hoff H., Heideberg J.,
RA Durkin A., Tatum M., Kallay T.E., Nelson M., Fedorova N.,
RA Grayam L.A., White O., Salzberg S.D., Lewis M.B., Kadane D.,
RA Holtzapple E., Abouli H., Wolf A.M., Ullrich T.E., Hamon C.L.,
RA M. J. J. J., Hoff H., Hoff H., Hoff H., Hoff H., Hoff H.,
RA Holt J.E., Loftus R.D., Yang F., Smith H.O., Venter J.C.,
RA Dougherty R.A., Hollnagel O.P., Fraser J.M.,
RA "Complete genome sequence of a wild-type isolate of Streptococcus
RT pneumoniae."
RI Science 297:404-409 (2002).
DR EMBL, AF007522, AF007522.1,
DR TIGR, ST0077, .
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 64 AA; 7330 MW; AC6B35D798041CF4 CRC64;

Query Match 66.7%; Score 4; DB 16; Length 64;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 3 RWHF 6
DB 26 RWHF 59

RESULT 11
Q94CT2
ID Q94CT2 PRELIMINARY; FRT; 76 AA.
AC Q94CT2;
DT 01-DEC-2001 (TREMBL) 19, Created;
DT 01-DEC-2001 (TREMBL) 19, Last sequence update;
DT 01-DEC-2001 (TREMBL) 19, Last annotation update;
DE P0459B04.14 protein.
GN P0459B04.14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Embryophyta;
OC Spermatophytes; Magnoliopsida; Liliaceae; Liliaceae;
OC Liliaceae; Liliaceae; Liliaceae; Liliaceae; Liliaceae;
OX NCBI_TaxID:4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV; NIPONBARE;
RA Sakai T., Matsumoto T., Yamamoto K.,
RA "Oryza sativa Nipponbare (CV) genome map.";
RT Genome 10:459B04.14;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RF EMBL, AF007522, AF007522.1, .
SQ SEQUENCE 76 AA; 8274 MW; 155AF5C83CD38BD CRC64;

Query Match 66.7%; Score 4; DB 10; Length 76;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 3 RWHF 6
DB 15 RWHF 18

RESULT 12
Q99LX1
ID Q99LX1 PRELIMINARY; FRT; 63 AA.
AC Q99LX1;
DT 01-JUN-2001 (TREMBL) 17, Created;
DT 01-JUN-2001 (TREMBL) 17, Last sequence update;
DT 01-DEC-2001 (TREMBL) 19, Last annotation update;
DE Hypothetical protein TFE2763.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mus musculus; Mus musculus; Mus musculus;
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SV; SV;
RA Straussberg R.,
RA Submitted (MAY 2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC002195, AAC02195.1, .

KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9486 MW; 509079R81C8941ED CRC64;

Query Match 66.7%; Score 4; DB 11; Length 83;
Best Local Similarity 100.0%; Pred No 2 99.02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 VWRW 4
|||
Db 54 VWRW 57

RESULT 13
Q96812 PRELIMINARY; PRT; 91 AA.

AC Q96812;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 10, Last annotation update)
DE RNA binding protein.
OS Grapevine virus D.
OC Viruses, ssRNA positive strand viruses, the DNA stage, Vitivirus
OX NCBI_TaxID=41617;
RN [1]
RP SEQUENCE FROM N.A.
RA Abou Ghannem N., Saldarelli P., Minafra A., Buzkan N., Castelli M A.,
RA Martelli G.P.;
RT "Properties of grapevine virus D, a novel putative 'Ritchovirus';"
RL Plant Pathol 78:15-25(1997)
DR EMBL; Y07764; CAA69071.1; -;
SQ SEQUENCE 91 AA; 10450 MW; 73F82F39F0248E8 CRC64;

Query Match 66.7%; Score 4; DB 12; Length 91;
Best Local Similarity 100.0%; Pred No 2.79e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 VWRW 4
|||
Db 70 VWRW 73

RESULT 14
Q8U0Y1 PRELIMINARY; PRT; 94 AA.

AC Q8U0Y1;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein PF1450.
GN PF1450.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=V01 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D M., Robb F T., Brown J P.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE010247; AAL81574.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 94 AA; 10554 MW; CE8AD48A76AF7CE8 CRC64;

Query Match 66.7%; Score 4; DB 17; Length 94;
Best Local Similarity 100.0%; Pred No 2 99.02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VWRW 4
|||
Db 91 VWRW 94

RESULT 15
Q9U8F9 PRELIMINARY; PRT; 98 AA.

AC Q9U8F9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein Atu4133.
GN Atu4133.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970);
OC Bacteria; Proteobacteria; Alpha subdivision; Rhizobiales group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE-21609550, PubMed-11743103,
PA Wood D W., Setubal J C., Paul P., Monte D E., Kitajima J.P.,
PA Okura V.K., Zhou Y., Chen L., Wood G E., Almeida N F. Jr., Wu L.,
PA Chen Y., Paulsen I T., Eison J A., Karp P E., Fernald D. Sr.,
PA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant G.,
PA Kutayavin T., Levy R., Li M.-J., McCallum E., Palmieri A., Perrin D.,
PA Paygood C., Pouse G., Sanger-Hamblin C., W. C., Fernald J., Perrin D.,
PA Zhang G., Yoo H., Tao Y., Riddle P., Jung M., Krogan W., Perry M.,
PA Gordon-Kamm B., Liao L., Kim S., Hendrick G., Zhao Z. Y., Delian M.,
PA Chumley F., Tingey S V., Tomb J.-F., Gordon M P., Olson M.V.,
PA Nester E.W.;
FT "The genome of the natural genetic engineer Agrobacterium tumefaciens
ET C58."
FT Science 294:2317-2322(2001).
PT EMBL; AE009344; AAL44933.1; -;
DR EMBL; AE009344; AAL44933.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 98 AA; 10925 MW; 830226F594293018 CRC64;

Query Match 66.7%; Score 4; DB 16; Length 98;
Best Local Similarity 100.0%; Pred No 2 99.02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VRWH 5
|||
Db 3 VRWH 6

Search completed: February 20, 2003, 11:28:56
Job time : 47 secs

GenCore version 5.1.3
Copyright 1993-2003 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: February 20, 2003, 10:35:40, Search time 23.333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-2

Perfect score: 45

Sequence: 1 WVPWHP 6

Scoring table: BIOSIMEX
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 131250620 residues

Total number of hits satisfying chosen parameters: 600470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: A.Geneseq 101002.*

1:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT*
2:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT*
3:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT*
4:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT*
5:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT*
6:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT*
7:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT*
8:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT*
9:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT*
10:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT*
11:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT*
12:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT*
13:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT*
14:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT*
15:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT*
16:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT*
17:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT*
18:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT*
19:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT*
20:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT*
21:	/SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT*
22:	/SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT*
23:	/SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	6	21	AAB01492
2	41	91.1	6	21	AAB01505
3	41	91.1	256	21	AAG09463
4	41	91.1	256	21	AAG44075
5	41	91.1	256	21	AAG48293
6	41	91.1	325	21	AAG09462
7	41	91.1	325	21	AAG44074
8	41	91.1	325	21	AAG48292
9	40	88.9	306	21	AAG05810
10	40	88.9	306	21	AAG47194

11	40	88.9	338	21	AAG05809	Arabidopsis thalia
12	40	88.9	338	21	AAG47193	Arabidopsis thalia
13	40	88.9	418	15	AAB0501	linoleic acid desat
14	40	88.9	448	21	AAG05808	Arabidopsis thalia
15	40	88.9	448	21	AAG47192	Arabidopsis thalia
16	40	88.9	448	23	ABP22179	Herbivore activity
17	39	86.7	6	21	AAB01506	peptide which bind
18	39	86.7	6	21	AAB01506	peptide which bind
19	38	84.4	477	21	AAV51600	Human truncated tr
20	38	84.4	822	16	AAB01630	Human trkB recepto
21	38	84.4	822	21	AAV51599	Human trkB recepto
22	38	84.4	822	23	AAV50951	Human receptor tyr
23	38	84.4	79	21	AAB40077	Human receptor tyr
24	36	80.0	79	23	ABP12225	Human GEF1 p
25	36	80.0	83	20	AAV25731	Human secreted pro
26	36	80.0	91	23	AP007463	Novel human diapho
27	36	80.0	424	22	AAE13431	Arabidopsis thaliana
28	36	80.0	1169	22	ABP1663	Drosophila melanog
29	35	77.8	6	21	AAB01508	Peptide which bind
30	35	77.8	53	23	ABP44902	Human albumin fusi
31	35	77.8	53	23	AAE21232	Human gene 18 caco
32	35	77.8	54	22	AAO09345	Human polypeptide
33	35	77.8	148	22	ABP11414	Human breast cancer
34	35	77.8	173	22	AAG80884	lipid modification
35	35	77.8	173	22	AAG83334	p-patent lipid met
36	35	77.8	663	22	AAU51530	Propionibacterium
37	34	75.6	6	21	AAB01504	peptide which bind
38	34	75.6	8	15	AAB60429	Antiproliferative
39	34	75.6	8	15	AAB60444	Antiproliferative
40	34	75.6	24	21	AAB58827	Human colon cancer
41	34	75.6	43	22	AAU14582	Novel bone marrow
42	34	75.6	49	22	AAU11182	Human leucocyte/haem
43	34	75.6	80	22	AAU55613	Human leucocyte/haem
44	34	75.6	36	22	AAU14777	Novel bone marrow
45	34	75.6	179	22	AAU44711	Propionibacterium

ALIGNMENTS

RESULT 1
AAB01492
ID AAB01492 standard; peptide; 6 AA.

AC AAB01492;
XX 08-NCV-2000 (first entry)
XX
XX
XX Peptide which binds to transcription factor E2F 1 DNA binding domain.
DE
DE
XX
XX
XX DNA binding, transcription factor, E2F, E2F 1, cell cycle; DP 1;
KW activation, transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
XX Synthetic.
OS
OS
XX
XX
XX WO200044771-A1.
XX
XX
XX 03-AUG-2000.
XX
XX 26-JAN-2000, 2000CW3 3900227.
XX
XX 26-JAN-2000, 2000CW3 3900227.
XX
XX 26-JAN-1999; 99GB-0001710.
XX
XX (PROL-) PROLIFIX LTD.
XX
XX Mueller R, Kontermann RE, Montigiani S;
XX
XX WFI; 2000-532806/48.
XX
XX Peptides binding to the FNA binding domain of transcription factor E2F
XX and inhibiting cell cycle progression, useful for the treatment of
XX cancer

XX Claim 6; Page 2; 42pp; English.
 XX
 CC Peptides which bind to the DNA binding domain of transcription
 CC factor E2F and inhibit cell cycle progression may be useful as
 CC research agents to investigate the interaction between E2F and DP-1,
 CC or the activation of transcription by E2F-1. Research agents may
 CC may also be used for inhibiting cell cycle arrest in
 CC a cell, particularly for treatment of cancer or other proliferative
 CC disorders such as perlestin and retinoblastoma.
 XX
 SC Sequence 5 AA;
 Query Match 100.0% Score 45; DB 21; Length 6;
 Best Local Similarity 100.0% Best Hit 100.0%
 Match 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WVRWHF 6
 Db 1 WVRWHF 6
 RESULT 2
 AAB01505
 ID AAB01505 standard; Peptide; 6 AA;
 XX
 AC AAB01505;
 XX
 DI 06 NOV-2000 (first entry)
 XX
 DE Peptide which binds to transcription factor E2F 2 DNA binding domain.
 XX
 KW DNA binding, transcription factor E2F 2, E2F 2, DNA binding domain,
 KW activation, transcription, apoptosis, proliferative disorder;
 KW proinflammation, neoplasia;
 XX
 CS Synthesis
 XX
 PN WC230044771-A1;
 XX
 PD 03 AUG-2000;
 XX
 PE 26 JAN-2001; 2000WV; 2000WV;
 XX
 FR 26 JAN-2001; 2000WV; 2000WV;
 XX
 PA 1902-1; 1902-1; 1902-1;
 XX
 PI Mueller R, Kottmann RE, Montgiani S,
 XX
 DR WP1; 2000-532806/48;
 XX
 PT Peptides binding to the DNA binding domain of transcription factor E2F
 PT and inhibiting cell cycle progression, useful for the treatment of
 PT cancer
 XX
 PS Example; Page 26; 42pp; English.
 XX
 CC Peptides which bind to the DNA binding domain of transcription
 CC factor E2F and inhibit cell cycle progression may be useful as
 CC research agents to investigate the interaction between E2F and DP-1,
 CC or the activation of transcription by E2F-1. Research agents may
 CC may also be used for inhibiting cell cycle arrest in
 CC a cell, particularly for treatment of cancer or other proliferative
 CC disorders such as perlestin and retinoblastoma.
 XX
 SC Sequence 5 AA;
 Query Match 91.1% Score 41; DB 21; Length 6;
 Best Local Similarity 91.1% Best Hit 100.0%
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 WVRWHF 6

Db 1 WVRWHF 6
 RESULT 3
 AAG09463
 ID AAG09463 standard; Peptide; 256 AA;
 XX
 AC AAG09463;
 XX
 DI 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment CFC ID NO 7409.
 XX
 KW Protein, Arabidopsis thaliana, signal transduction pathway,
 KW hybridization assay, gene expression, gene expression control,
 KW termination sequence.
 XX
 CS Arabidopsis thaliana
 XX
 PN EP1033405-A2;
 XX
 PD 06-SEP-2000;
 XX
 PE 25-FEB-1999; 2000EP-030439;
 XX
 DE 25-FEB-1999; 99US-0121825;
 PR 05-MAR-1999; 99US-012180;
 PR 05-MAR-1999; 99US-0123548;
 PR 23-MAR-1999; 99US-0125788;
 PR 05-MAR-1999; 99US-0126264;
 PR 29-MAR-1999; 99US-0126785;
 PR 01-MAR-1999; 99US-0127362;
 PR 06-APR-1999; 99US-0128234;
 PR 06-APR-1999; 99US-0128714;
 PR 16-APR-1999; 99US-0129845;
 PR 19-APR-1999; 99US-0130077;
 PR 21-APR-1999; 99US-0130449;
 PR 23-APR-1999; 99US-0130510;
 PR 23-APR-1999; 99US-0130891;
 PR 28-APR-1999; 99US-0131449;
 PR 30-APR-1999; 99US-0132048;
 PR 30-APR-1999; 99US-0132407;
 PR 04-MAY-1999; 99US-0132484;
 PR 05-MAY-1999; 99US-0132485;
 PR 06-MAY-1999; 99US-0132486;
 PR 06-MAY-1999; 99US-0132487;
 PR 07-MAY-1999; 99US-0132488;
 PR 11-MAY-1999; 99US-0132506;
 PR 14-MAY-1999; 99US-0134218;
 PR 14-MAY-1999; 99US-0134219;
 PR 14-MAY-1999; 99US-0134221;
 PR 14-MAY-1999; 99US-0134370;
 PR 19-MAY-1999; 99US-0134749;
 PR 19-MAY-1999; 99US-0134941;
 PR 20-MAY-1999; 99US-0135124;
 PR 21-MAY-1999; 99US-0135353;
 PR 24-MAY-1999; 99US-0135629;
 PR 25-MAY-1999; 99US-0136021;
 PR 27-MAY-1999; 99US-0136302;
 PR 28-MAY-1999; 99US-0136782;
 PR 01-JUN-1999; 99US-0137002;
 PR 11-JUN-1999; 99US-0137028;
 PR 14-JUN-1999; 99US-0137500;
 PR 07-JUN-1999; 99US-0137724;
 PR 08-JUN-1999; 99US-0138094;
 PR 10-JUN-1999; 99US-0138540;
 PR 10-JUN-1999; 99US-0138847;
 PR 14-JUN-1999; 99US-0139119;
 PR 16-JUN-1999; 99US-0139453;
 PR 17-JUN-1999; 99US-0139492;
 PR 18-JUN-1999; 99US-0139494;

[illegible]

PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	22-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149829
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	09-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361
PR	28-OCT-1999;	99US-0161320
PR	28-OCT-1999;	99US-0161992
PR	28-OCT-1999;	99US-0161993
PR	29-OCT-1999;	99US-0162142

RESULT 4
AAG44075
ID AAG44075 standard; Protein; 256 AA.

CY	1	WVRWHF	6
Db	72	WARWHF	77

```

Query Match      91.1%; Score 41; DB 21; Length 256;
Best Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 WVRWHF 6

Db 72 WARWHE 77

AA644075

ID AAG44075 standard; Protein; 256 AA.

[illegible]

PK	18-JUN-1999;	9908-0139462;
PR	18-JUN-1999;	9908-0139473;
PK	18-JUN-1999;	9908-0139750;
PK	18-JUN-1999;	9908-0139763;
PR	21-JUN-1999;	9908-0139817;
PR	21-JUN-1999;	9908-0139849;
PR	23-JUN-1999;	9908-0140353;
PR	23-JUN-1999;	9908-0140354;
PR	24-JUN-1999;	9908-0140395;
PK	28-JUN-1999;	9908-0140823;
PK	29-JUN-1999;	9908-0140991;
PR	30-JUN-1999;	9908-0141287;
PR	01-JUL-1999;	9908-0141842;
PR	01-JUL-1999;	9908-0142154;
PR	02-JUL-1999;	9908-0142055;
PK	06-JUL-1999;	9908-0142390;
PK	08-JUL-1999;	9908-0142803;
PR	09-JUL-1999;	9908-0142920;
PK	12-JUL-1999;	9908-0142977;
PK	13-JUL-1999;	9908-0143542;
PK	14-JUL-1999;	9908-0143624;
PK	15-JUL-1999;	9908-0144005;
PK	16-JUL-1999;	9908-0144085;
PK	16-JUL-1999;	9908-0144096;
PK	18-JUL-1999;	9908-0144325;
PR	19-JUL-1999;	9908-0144331;
PR	19-JUL-1999;	9908-0144332;
PR	19-JUL-1999;	9908-0144333;
PR	19-JUL-1999;	9908-0144334;
PR	19-JUL-1999;	9908-0144335;
PR	20-JUL-1999;	9908-0144336;
PR	20-JUL-1999;	9908-0144337;
PR	20-JUL-1999;	9908-0144338;
PR	21-JUL-1999;	9908-0144339;
PR	21-JUL-1999;	9908-0144340;
PR	21-JUL-1999;	9908-0144341;
PR	21-JUL-1999;	9908-0144342;
PR	21-JUL-1999;	9908-0144343;
PR	21-JUL-1999;	9908-0144344;
PR	21-JUL-1999;	9908-0144345;
PR	21-JUL-1999;	9908-0144346;
PR	21-JUL-1999;	9908-0144347;
PR	21-JUL-1999;	9908-0144348;
PR	21-JUL-1999;	9908-0144349;
PR	21-JUL-1999;	9908-0144350;
PR	21-JUL-1999;	9908-0144351;
PR	21-JUL-1999;	9908-0144352;
PR	21-JUL-1999;	9908-0144353;
PR	21-JUL-1999;	9908-0144354;
PR	21-JUL-1999;	9908-0144355;
PR	21-JUL-1999;	9908-0144356;
PR	21-JUL-1999;	9908-0144357;
PR	21-JUL-1999;	9908-0144358;
PR	21-JUL-1999;	9908-0144359;
PR	21-JUL-1999;	9908-0144360;
PR	21-JUL-1999;	9908-0144361;
PR	21-JUL-1999;	9908-0144362;
PR	21-JUL-1999;	9908-0144363;
PR	21-JUL-1999;	9908-0144364;
PR	21-JUL-1999;	9908-0144365;
PR	21-JUL-1999;	9908-0144366;
PR	21-JUL-1999;	9908-0144367;
PR	21-JUL-1999;	9908-0144368;
PR	21-JUL-1999;	9908-0144369;
PR	21-JUL-1999;	9908-0144370;
PR	21-JUL-1999;	9908-0144371;
PR	21-JUL-1999;	9908-0144372;
PR	21-JUL-1999;	9908-0144373;
PR	21-JUL-1999;	9908-0144374;
PR	21-JUL-1999;	9908-0144375;
PR	21-JUL-1999;	9908-0144376;
PR	21-JUL-1999;	9908-0144377;
PR	21-JUL-1999;	9908-0144378;
PR	21-JUL-1999;	9908-0144379;
PR	21-JUL-1999;	9908-0144380;
PR	21-JUL-1999;	9908-0144381;
PR	21-JUL-1999;	9908-0144382;
PR	21-JUL-1999;	9908-0144383;
PR	21-JUL-1999;	9908-0144384;
PR	21-JUL-1999;	9908-0144385;
PR	21-JUL-1999;	9908-0144386;
PR	21-JUL-1999;	9908-0144387;
PR	21-JUL-1999;	9908-0144388;
PR	21-JUL-1999;	9908-0144389;
PR	21-JUL-1999;	9908-0144390;
PR	21-JUL-1999;	9908-0144391;
PR	21-JUL-1999;	9908-0144392;
PR	21-JUL-1999;	9908-0144393;
PR	21-JUL-1999;	9908-0144394;
PR	21-JUL-1999;	9908-0144395;
PR	21-JUL-1999;	9908-0144396;
PR	21-JUL-1999;	9908-0144397;
PR	21-JUL-1999;	9908-0144398;
PR	21-JUL-1999;	9908-0144399;
PR	21-JUL-1999;	9908-0144400;
PR	21-JUL-1999;	9908-0144401;
PR	21-JUL-1999;	9908-0144402;
PR	21-JUL-1999;	9908-0144403;
PR	21-JUL-1999;	9908-0144404;
PR	21-JUL-1999;	9908-014440

PR	25	AUG-1999;	99US-0150866
PR	26	AUG-1999;	99US-0150884
PR	27	AUG-1999;	99US-0151065
PR	27	AUG-1999;	99US-0151066
PR	27	AUG-1999;	99US-0151080
PR	30	AUG-1999;	99US-0151303
PR	31	AUG-1999;	99US-0151438
PR	01	SEP-1999;	99US-0151910
PR	07	SEP-1999;	99US-0152363
PR	10	SEP-1999;	99US-0153070
PR	13	SEP-1999;	99US-0153758
PR	15	SEP-1999;	99US-0154018
PR	16	SEP-1999;	99US-0154019
PR	20	SEP-1999;	99US-0154779
PR	22	SEP-1999;	99US-0155139
PR	23	SEP-1999;	99US-0155486
PR	24	SEP-1999;	99US-0155659
PR	28	SEP-1999;	99US-0156459
PR	29	SEP-1999;	99US-0156596
PR	04	OCT-1999;	99US-0157117
PR	05	OCT-1999;	99US-0157753
PR	06	OCT-1999;	99US-0157865
PR	07	OCT-1999;	99US-0158029
PR	08	OCT-1999;	99US-0158232
PR	12	OCT-1999;	99US-0158369
PR	13	OCT-1999;	99US-0159294
PR	13	OCT-1999;	99US-0159295
PR	14	OCT-1999;	99US-0159329
PR	14	OCT-1999;	99US-0159330
PR	14	OCT-1999;	99US-0159331
PR	14	OCT-1999;	99US-0159637
PR	14	OCT-1999;	99US-0159638
PR	18	OCT-1999;	99US-0159584
PR	21	OCT-1999;	99US-0160741
PR	21	OCT-1999;	99US-0160767
PR	21	OCT-1999;	99US-0160768
PR	21	OCT-1999;	99US-0160770
PR	21	OCT-1999;	99US-0160814
PR	21	OCT-1999;	99US-0160815
PR	22	OCT-1999;	99US-0160980
PR	22	OCT-1999;	99US-0160981
PR	22	OCT-1999;	99US-0160989
PR	25	OCT-1999;	99US-0161404
PR	25	OCT-1999;	99US-0161405
PR	25	OCT-1999;	99US-0161406
PR	26	OCT-1999;	99US-0161359
PR	26	OCT-1999;	99US-0161360
PR	26	OCT-1999;	99US-0161361
PR	28	OCT-1999;	99US-0161920
PR	28	OCT-1999;	99US-0161992
PR	28	OCT-1999;	99US-0161993
PR	29	OCT-1999;	99US-0162142

Query Match 91.1%; Score 41; DB 21; Length 256;
 Best Local Similarity 83.3%; Pred. No 35;
 Matches 5; Conservative 0; Mismatches 1; Indels 0, Gaps 0,

QY 1 WVRWHF 6
 Db 72 WVRWHF 77
 RESULT 5
 AAG48293 standard; Protein; 256 AA.
 XX
 AC AAG48293;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO. 60969
 XX

KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
FN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PP	25-FEB-1999; 99US-0121825
PP	05-MAR-1999; 99US-0123180
PP	09-MAR-1999; 99US-0123548
PP	23-MAR-1999; 99US-0125788
PP	25-MAR-1999; 99US-0126264
PP	29-MAR-1999; 99US-0126785
PP	01-APR-1999; 99US-0127462
PP	06-APR-1999; 99US-0128234
PP	08-APR-1999; 99US-0128714
PP	16-APR-1999; 99US-0129845
PP	19-APR-1999; 99US-0130077
PP	21-APR-1999; 99US-0130449
PP	23-APR-1999; 99US-0130510
PP	28-APR-1999; 99US-0130891
PP	30-APR-1999; 99US-0131449
PP	30-APR-1999; 99US-0132048
PP	30-APR-1999; 99US-0132407
PP	04-MAY-1999; 99US-0132484
PP	05-MAY-1999; 99US-0132485
PP	06-MAY-1999; 99US-0132486
PP	06-MAY-1999; 99US-0132487
PP	07-MAY-1999; 99US-0132863
PP	11-MAY-1999; 99US-0134256
PP	14-MAY-1999; 99US-0134218
PP	14-MAY-1999; 99US-0134219
PP	14-MAY-1999; 99US-0134221
PP	14-MAY-1999; 99US-0134370
PP	18-MAY-1999; 99US-0134768
PP	19-MAY-1999; 99US-0134941
PP	20-MAY-1999; 99US-0135124
PP	21-MAY-1999; 99US-0135353
PP	24-MAY-1999; 99US-0135629
PP	25-MAY-1999; 99US-0136021
PP	27-MAY-1999; 99US-0136392
PP	29-MAY-1999; 99US-0136782
PP	01-JUN-1999; 99US-0137222
PP	03-JUN-1999; 99US-0137528
PP	04-JUN-1999; 99US-0137502
PP	07-JUN-1999; 99US-0137724
PP	08-JUN-1999; 99US-0138094
PP	10-JUN-1999; 99US-0138540
PP	10-JUN-1999; 99US-0138847
PP	14-JUN-1999; 99US-0139119
PP	16-JUN-1999; 99US-0139452
PP	16-JUN-1999; 99US-0139453
PP	17-JUN-1999; 99US-0139492
PP	18-JUN-1999; 99US-0139454
PP	18-JUN-1999; 99US-0139455
PP	18-JUN-1999; 99US-0139456
PP	18-JUN-1999; 99US-0139457
PP	18-JUN-1999; 99US-0139458
PP	18-JUN-1999; 99US-0139459
PP	18-JUN-1999; 99US-0139460
PP	18-JUN-1999; 99US-0139461
PP	18-JUN-1999; 99US-0139462
PP	18-JUN-1999; 99US-0139463
PP	18-JUN-1999; 99US-0139750
PP	18-JUN-1999; 99US-0139763
PP	21-JUN-1999; 99US-0139817
PP	22-JUN-1999; 99US-0139899
PP	23-JUN-1999; 99US-0140353

PR	23-JUN-1999;	9905-0140354
PR	24-JUN-1999;	9905-0140365
PR	28-JUN-1999;	9905-0140823
PR	29-JUN-1999;	9905-0140991
PR	30-JUN-1999;	9905-0141297
PR	01-JUL-1999;	9905-0141942
PR	02-JUL-1999;	9905-0141954
PR	06-JUL-1999;	9905-0141996
PR	08-JUL-1999;	9905-0142903
PR	10-JUL-1999;	9905-0143420
PR	13-JUL-1999;	9905-0144257
PR	14-JUL-1999;	9905-0144542
PR	15-JUL-1999;	9905-0144624
PR	16-JUL-1999;	9905-0144805
PR	16-JUL-1999;	9905-0144956
PR	19-JUL-1999;	9905-0144931
PR	19-JUL-1999;	9905-0144932
PR	19-JUL-1999;	9905-0144933
PR	19-JUL-1999;	9905-0144934
PR	19-JUL-1999;	9905-0144935
PR	20-JUL-1999;	9905-0144936
PR	20-JUL-1999;	9905-0144937
PR	21-JUL-1999;	9905-0144981
PR	21-JUL-1999;	9905-0145086
PR	22-JUL-1999;	9905-0145085
PR	22-JUL-1999;	9905-0145087
PR	22-JUL-1999;	9905-0145089
PR	23-JUL-1999;	9905-0145090
PR	23-JUL-1999;	9905-0145091
PR	23-JUL-1999;	9905-0145092
PR	23-JUL-1999;	9905-0145093
PR	23-JUL-1999;	9905-0145094
PR	23-JUL-1999;	9905-0145095
PR	23-JUL-1999;	9905-0145096
PR	23-JUL-1999;	9905-0145097
PR	23-JUL-1999;	9905-0145098
PR	23-JUL-1999;	9905-0145099
PR	23-JUL-1999;	9905-0145100
PR	23-JUL-1999;	9905-0145101
PR	23-JUL-1999;	9905-0145102
PR	23-JUL-1999;	9905-0145103
PR	23-JUL-1999;	9905-0145104
PR	23-JUL-1999;	9905-0145105
PR	23-JUL-1999;	9905-0145106
PR	23-JUL-1999;	9905-0145107
PR	23-JUL-1999;	9905-0145108
PR	23-JUL-1999;	9905-0145109
PR	23-JUL-1999;	9905-0145110
PR	23-JUL-1999;	9905-0145111
PR	23-JUL-1999;	9905-0145112
PR	23-JUL-1999;	9905-0145113
PR	23-JUL-1999;	9905-0145114
PR	23-JUL-1999;	9905-0145115
PR	23-JUL-1999;	9905-0145116
PR	23-JUL-1999;	9905-0145117
PR	23-JUL-1999;	9905-0145118
PR	23-JUL-1999;	9905-0145119
PR	23-JUL-1999;	9905-0145120
PR	23-JUL-1999;	9905-0145121
PR	23-JUL-1999;	9905-0145122
PR	23-JUL-1999;	9905-0145123
PR	23-JUL-1999;	9905-0145124
PR	23-JUL-1999;	9905-0145125
PR	23-JUL-1999;	9905-0145126
PR	23-JUL-1999;	9905-0145127
PR	23-JUL-1999;	9905-0145128
PR	23-JUL-1999;	9905-0145129
PR	23-JUL-1999;	9905-0145130
PR	23-JUL-1999;	9905-0145131
PR	23-JUL-1999;	9905-0145132
PR	23-JUL-1999;	9905-0145133
PR	23-JUL-1999;	9905-0145134
PR	23-JUL-1999;	9905-0145135
PR	23-JUL-1999;	9905-0145136
PR	23-JUL-1999;	9905-0145137
PR	23-JUL-1999;	9905-0145138
PR	23-JUL-1999;	9905-0145139
PR	23-JUL-1999;	9905-0145140
PR	23-JUL-1999;	9905-0145141
PR	23-JUL-1999;	9905-0145142
PR	23-JUL-1999;	9905-0145143
PR	23-JUL-1999;	9905-0145144
PR	23-JUL-1999;	9905-0145145
PR	23-JUL-1999;	9905-0145146
PR	23-JUL-1999;	9905-0145147
PR	23-JUL-1999;	9905-0145148
PR	23-JUL-1999;	9905-0145149
PR	23-JUL-1999;	9905-0145150
PR	23-JUL-1999;	9905-0145151
PR	23-JUL-1999;	9905-0145152
PR	23-JUL-1999;	9905-0145153
PR	23-JUL-1999;	9905-0145154
PR	23-JUL-1999;	9905-0145155
PR	23-JUL-1999;	9905-0145156
PR	23-JUL-1999;	9905-0145157
PR	23-JUL-1999;	9905-0145158
PR	23-JUL-1999;	9905-0145159
PR	23-JUL-1999;	9905-0145160
PR	23-JUL-1999;	9905-0145161
PR	23-JUL-1999;	9905-014516

PR	07-SEP-1999	9905-01519330
PR	07-SEP-1999	9905-01522633
PR	10-SEP-1999	9905-01533070
PR	12-SEP-1999	9905-01537588
PR	15-SEP-1999	9905-01540418
PR	16-SEP-1999	9905-01540399
PR	20-SEP-1999	9905-01540779
PR	22-SEP-1999	9905-01551390
PR	22-SEP-1999	9905-01551466
PR	24-SEP-1999	9905-01556659
PR	24-SEP-1999	9905-01564498
PR	25-SEP-1999	9905-01566996
PR	04-OCT-1999	9905-01571117
PR	05-OCT-1999	9905-01577553
PR	06-OCT-1999	9905-01578655
PR	07-OCT-1999	9905-01580629
PR	08-OCT-1999	9905-01586232
PR	12-OCT-1999	9905-01584293
PR	13-OCT-1999	9905-01592293
PR	13-OCT-1999	9905-01592994
PR	13-OCT-1999	9905-01592995
PR	14-OCT-1999	9905-01593329
PR	14-OCT-1999	9905-01593330
PR	14-OCT-1999	9905-01595933
PR	14-OCT-1999	9905-01596837
PR	14-OCT-1999	9905-01596838
PR	18-OCT-1999	9905-01597584
PR	21-OCT-1999	9905-01600941
PR	21-OCT-1999	9905-01600967
PR	21-OCT-1999	9905-01607068
PR	21-OCT-1999	9905-01607770
PR	21-OCT-1999	9905-01608141
PR	21-OCT-1999	9905-01608142
PR	21-OCT-1999	9905-01609080
PR	22-OCT-1999	9905-01609981
PR	22-OCT-1999	9905-01610394
PR	23-OCT-1999	9905-01614055
PR	25-OCT-1999	9905-01614065
PR	25-OCT-1999	9905-01614066
PR	26-OCT-1999	9905-01613560
PR	26-OCT-1999	9905-01613611
PR	28-OCT-1999	9905-01619920
PR	28-OCT-1999	9905-01619921
PR	29-OCT-1999	9905-01623422

Category	Count	Percentage	Mean	Std. Dev.	Min.	Max.
1. No match	1	100.0%	1.00	.00	1	1
2. 1 match	0	0.0%	2.00	.00	2	2
3. 2 matches	0	0.0%	3.00	.00	3	3
4. 3 matches	0	0.0%	4.00	.00	4	4
5. 4 matches	0	0.0%	5.00	.00	5	5
6. 5 matches	0	0.0%	6.00	.00	6	6
7. 6 matches	0	0.0%	7.00	.00	7	7
8. 7 matches	0	0.0%	8.00	.00	8	8
9. 8 matches	0	0.0%	9.00	.00	9	9
10. 9 matches	0	0.0%	10.00	.00	10	10
11. 10 matches	0	0.0%	11.00	.00	11	11
12. 11 matches	0	0.0%	12.00	.00	12	12
13. 12 matches	0	0.0%	13.00	.00	13	13
14. 13 matches	0	0.0%	14.00	.00	14	14
15. 14 matches	0	0.0%	15.00	.00	15	15
16. 15 matches	0	0.0%	16.00	.00	16	16
17. 16 matches	0	0.0%	17.00	.00	17	17
18. 17 matches	0	0.0%	18.00	.00	18	18
19. 18 matches	0	0.0%	19.00	.00	19	19
20. 19 matches	0	0.0%	20.00	.00	20	20
21. 20 matches	0	0.0%	21.00	.00	21	21
22. 21 matches	0	0.0%	22.00	.00	22	22
23. 22 matches	0	0.0%	23.00	.00	23	23
24. 23 matches	0	0.0%	24.00	.00	24	24
25. 24 matches	0	0.0%	25.00	.00	25	25
26. 25 matches	0	0.0%	26.00	.00	26	26
27. 26 matches	0	0.0%	27.00	.00	27	27
28. 27 matches	0	0.0%	28.00	.00	28	28
29. 28 matches	0	0.0%	29.00	.00	29	29
30. 29 matches	0	0.0%	30.00	.00	30	30
31. 30 matches	0	0.0%	31.00	.00	31	31
32. 31 matches	0	0.0%	32.00	.00	32	32
33. 32 matches	0	0.0%	33.00	.00	33	33
34. 33 matches	0	0.0%	34.00	.00	34	34
35. 34 matches	0	0.0%	35.00	.00	35	35
36. 35 matches	0	0.0%	36.00	.00	36	36
37. 36 matches	0	0.0%	37.00	.00	37	37
38. 37 matches	0	0.0%	38.00	.00	38	38
39. 38 matches	0	0.0%	39.00	.00	39	39
40. 39 matches	0	0.0%	40.00	.00	40	40
41. 40 matches	0	0.0%	41.00	.00	41	41
42. 41 matches	0	0.0%	42.00	.00	42	42
43. 42 matches	0	0.0%	43.00	.00	43	43
44. 43 matches	0	0.0%	44.00	.00	44	44
45. 44 matches	0	0.0%	45.00	.00	45	45
46. 45 matches	0	0.0%	46.00	.00	46	46
47. 46 matches	0	0.0%	47.00	.00	47	47
48. 47 matches	0	0.0%	48.00	.00	48	48
49. 48 matches	0	0.0%	49.00	.00	49	49
50. 49 matches	0	0.0%	50.00	.00	50	50
51. 50 matches	0	0.0%	51.00	.00	51	51
52. 51 matches	0	0.0%	52.00	.00	52	52
53. 52 matches	0	0.0%	53.00	.00	53	53
54. 53 matches	0	0.0%	54.00	.00	54	54
55. 54 matches	0	0.0%	55.00	.00	55	55
56. 55 matches	0	0.0%	56.00	.00	56	56
57. 56 matches	0	0.0%	57.00	.00	57	57
58. 57 matches	0</					

QY	1	WYRWHF	6
		1 1 1 1	
DB	72	WARWHF	77

RESULT 6
AA509462
111 Amino acid standard; Protein; 325 AA

DT 17-OCT-2000 (first entry)

DE
ARLINGTON Hall and Protein fragment CPO IN NO. 7408

[illegible]

FM
terminating sequence.

Stefan Julius Hallama.

EP1033405-A2

XX 06-SEP-2000 2000FP-0301430
XX
PR 25-FEB-1999 99US-0121925
PR 05-MAR-1999 99US-0123180
PR 09-MAR-1999 99US-0123548
PR 23-MAR-1999 99US-0125799
PR 25-MAR-1999 99US-0126264
PR 29-MAR-1999 99US-0126785
PR 01-APR-1999 99US-0127462
PR 06-APR-1999 99US-0128234
PR 08-APR-1999 99US-0128714
PR 16-APR-1999 99US-0129845
PR 19-APR-1999 99US-0130077
PR 21-APR-1999 99US-0130449
PR 23-APR-1999 99US-0130510
PR 23-APR-1999 99US-0130891
PR 28-APR-1999 99US-0131449
PR 30-APR-1999 99US-0132048
PR 30-APR-1999 99US-0132407
PR 04-MAY-1999 99US-0132484
PR 05-MAY-1999 99US-0132485
PR 06-MAY-1999 99US-0132486
PR 06-MAY-1999 99US-0132487
PR 07-MAY-1999 99US-0132863
PR 11-MAY-1999 99US-0134256
PR 14-MAY-1999 99US-0134218
PR 14-MAY-1999 99US-0134219
PR 14-MAY-1999 99US-0134221
PR 14-MAY-1999 99US-0134370
PR 18-MAY-1999 99US-0134768
PR 19-MAY-1999 99US-0134941
PR 20-MAY-1999 99US-0135124
PR 21-MAY-1999 99US-0135353
PR 24-MAY-1999 99US-0135629
PR 25-MAY-1999 99US-0136021
PR 27-MAY-1999 99US-0136392
PR 28-MAY-1999 99US-0136782
PR 01-JUN-1999 99US-0137222
PR 03-JUN-1999 99US-0137529
PR 04-JUN-1999 99US-0137502
PR 07-JUN-1999 99US-0137724
PR 08-JUN-1999 99US-0139094
PR 10-JUN-1999 99US-0138540
PR 10-JUN-1999 99US-0138847
PR 14-JUN-1999 99US-0139119
PR 16-JUN-1999 99US-0139452
PR 16-JUN-1999 99US-0139453
PR 17-JUN-1999 99US-0139492
PR 18-JUN-1999 99US-0139454
PR 18-JUN-1999 99US-0139455
PR 18-JUN-1999 99US-0139456
PR 18-JUN-1999 99US-0139457
PR 18-JUN-1999 99US-0139458
PR 18-JUN-1999 99US-0139459
PR 18-JUN-1999 99US-0139460
PR 18-JUN-1999 99US-0139461
PR 18-JUN-1999 99US-0139462
PR 18-JUN-1999 99US-0139463
PR 18-JUN-1999 99US-0139750
PR 18-JUN-1999 99US-0139763
PR 21-JUN-1999 99US-0139817
PR 22-JUN-1999 99US-0139899
PR 23-JUN-1999 99US-0140353
PR 23-JUN-1999 99US-0140354
PR 24-JUN-1999 99US-0140696
PR 28-JUN-1999 99US-0140823
PR 29-JUN-1999 99US-0140991
PR 30-JUN-1999 99US-0141287
PR 01-JUL-1999 99US-0141842
PR 01-JUL-1999 99US-0142154

PR 02-JUL-1999 99US-0142055
PR 06-JUL-1999 99US-0142390
PR 08-JUL-1999 99US-0142803
PR 09-JUL-1999 99US-0142920
PR 12-JUL-1999 99US-0142977
PR 13-JUL-1999 99US-0143542
PR 14-JUL-1999 99US-0143624
PR 15-JUL-1999 99US-0144005
PR 16-JUL-1999 99US-0144085
PR 16-JUL-1999 99US-0144086
PR 19-JUL-1999 99US-0144326
PR 19-JUL-1999 99US-0144331
PR 19-JUL-1999 99US-0144332
PR 19-JUL-1999 99US-0144333
PR 19-JUL-1999 99US-0144334
PR 19-JUL-1999 99US-0144335
PR 20-JUL-1999 99US-0144352
PR 20-JUL-1999 99US-0144632
PR 20-JUL-1999 99US-0144884
PR 21-JUL-1999 99US-0144814
PR 21-JUL-1999 99US-0145086
PR 21-JUL-1999 99US-0145088
PR 22-JUL-1999 99US-0145089
PR 22-JUL-1999 99US-0145087
PR 22-JUL-1999 99US-0145089
PR 22-JUL-1999 99US-0145192
PR 23-JUL-1999 99US-0145145
PR 23-JUL-1999 99US-0145218
PR 23-JUL-1999 99US-0145224
PR 26-JUL-1999 99US-0145276
PR 27-JUL-1999 99US-0145913
PR 27-JUL-1999 99US-0145918
PR 27-JUL-1999 99US-0145919
PR 28-JUL-1999 99US-0145951
PR 02-AUG-1999 99US-0146386
PR 02-AUG-1999 99US-0146388
PR 02-AUG-1999 99US-0146389
PR 03-AUG-1999 99US-0147038
PR 04-AUG-1999 99US-0147204
PR 04-AUG-1999 99US-0147302
PR 05-AUG-1999 99US-0147192
PR 05-AUG-1999 99US-0147260
PR 06-AUG-1999 99US-0147303
PR 06-AUG-1999 99US-0147416
PR 09-AUG-1999 99US-0147493
PR 09-AUG-1999 99US-0147935
PR 10-AUG-1999 99US-0148171
PR 11-AUG-1999 99US-0148319
PR 12-AUG-1999 99US-0148341
PR 13-AUG-1999 99US-0148565
PR 13-AUG-1999 99US-0148684
PR 16-AUG-1999 99US-0149368
PR 17-AUG-1999 99US-0149175
PR 18-AUG-1999 99US-0149426
PR 20-AUG-1999 99US-0149722
PR 20-AUG-1999 99US-0149723
PR 20-AUG-1999 99US-0149929
PR 23-AUG-1999 99US-0149902
PR 23-AUG-1999 99US-0149930
PR 25-AUG-1999 99US-0150566
PR 26-AUG-1999 99US-0150884
PR 27-AUG-1999 99US-0151065
PR 27-AUG-1999 99US-0151065
PR 27-AUG-1999 99US-0151065
PR 27-AUG-1999 99US-0151080
PR 30-AUG-1999 99US-0151303
PR 31-AUG-1999 99US-0151438
PR 01-SEP-1999 99US-0151930
PR 07-SEP-1999 99US-0152363
PR 10-SEP-1999 99US-0153070
PR 13-SEP-1999 99US-0153758
PR 15-SEP-1999 99US-0154018
PR 16-SEP-1999 99US-0154039
PR 20-SEP-1999 99US-0154779

PR 42 SEP 1999; 9905-0155134;
PR 23-SEP-1999; 9905-0155484;
PR 24-SEP-1999; 9905-0155484;
PR 28 SEP 1999; 9905-0155456;
PR 28-SEP-1999; 9905-0155456;
PR 24-SEP-1999; 9905-0157117;
PR 05-OCT-1999; 9905-0157753;
PR 06-OCT-1999; 9905-0157885;
PR 07-OCT-1999; 9905-0158029;
PR 08-OCT-1999; 9905-0158232;
PR 12-OCT-1999; 9905-0158363;
PR 13-OCT-1999; 9905-0158453;
PR 13-OCT-1999; 9905-0159294;
PR 13-OCT-1999; 9905-0159455;
PR 14-OCT-1999; 9905-0159429;
PR 14-OCT-1999; 9905-0159330;
PR 14-OCT-1999; 9905-0159331;
PR 14-OCT-1999; 9905-0159447;
PR 14-OCT-1999; 9905-0159448;
PR 14-OCT-1999; 9905-0159884;
PR 21-OCT-1999; 9905-0160741;
PR 21-OCT-1999; 9905-0160767;
PR 21-OCT-1999; 9905-0160768;
PR 21-OCT-1999; 9905-0160770;
PR 21-OCT-1999; 9905-0160814;
PR 21-OCT-1999; 9905-0160815;
PR 21-OCT-1999; 9905-0160880;
PR 22-OCT-1999; 9905-0160881;
PR 22-OCT-1999; 9905-0160989;
PR 23-OCT-1999; 9905-0161404;
PR 25-OCT-1999; 9905-0161405;
PR 25-OCT-1999; 9905-0161406;
PR 25-OCT-1999; 9905-0161359;
PR 26-OCT-1999; 9905-0161360;
PR 26-OCT-1999; 9905-0161361;
PR 26-OCT-1999; 9905-0161361;
PR 28-OCT-1999; 9905-0161992;
PR 28-OCT-1999; 9905-0161993;
PR 29-OCT-1999; 9905-0162142;

Query Match 91.1%; Score 41; E 21; Length 325;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 37 Conservation 27; Index 37; gaps 0;

OY : WVRWHP 6
DB 141 WVRWHP 146

RESULT 7
AAG44074
ID AAG44074 standard, Protein, 325 aa.
XX
AC AAG44074;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arbidopois chlamydia protein fragment SEQ ID NO: 5510;
XX
KW Protein identification, signal transduction pathway, metabolic pathway,
KW hydrolysis assay, genetic, amino acid, gene expression, cell biology,
KW termination sequence.
XX
OS Arbidopois thalanda
XX
PN EF1033405-A2
XX
ED 06-DEC-2000;
XX
PF 25-FEB-2000; 2000EF-0101449
XX
PR 25-FEB-1999; 9905-0121825;
PR 25-MAR-1999; 9905-0123180;

PR 09-MAR-1999; 9905-0123548;
PR 23-MAR-1999; 9905-0125788;
PR 25-MAR-1999; 9905-0125864;
PR 23-MAR-1999; 9905-0126785;
PR 01-APR-1999; 9905-0127462;
PR 06-APR-1999; 9905-0128234;
PR 08-APR-1999; 9905-0128714;
PR 16-APR-1999; 9905-0128845;
PR 19-APR-1999; 9905-0130077;
PR 21-APR-1999; 9905-0130449;
PR 23-APR-1999; 9905-0130510;
PR 23-APR-1999; 9905-0130891;
PR 28-APR-1999; 9905-0131449;
PR 30-APR-1999; 9905-0132048;
PR 30-APR-1999; 9905-0132407;
PR 04-MAY-1999; 9905-0132484;
PR 05-MAY-1999; 9905-0132485;
PR 06-MAY-1999; 9905-0132486;
PR 06-MAY-1999; 9905-0132487;
PR 07-MAY-1999; 9905-0132863;
PR 11-MAY-1999; 9905-0134256;
PR 14-MAY-1999; 9905-0134218;
PR 14-MAY-1999; 9905-0134219;
PR 14-MAY-1999; 9905-0134221;
PR 14-MAY-1999; 9905-0134370;
PR 18-MAY-1999; 9905-0134768;
PR 19-MAY-1999; 9905-0134941;
PR 20-MAY-1999; 9905-0135124;
PR 21-MAY-1999; 9905-0135353;
PR 24-MAY-1999; 9905-0135629;
PR 25-MAY-1999; 9905-0135621;
PR 27-MAY-1999; 9905-0135639;
PR 28-MAY-1999; 9905-0135692;
PR 01-JUN-1999; 9905-0137222;
PR 03-JUN-1999; 9905-0137128;
PR 04-JUN-1999; 9905-0137502;
PR 07-JUN-1999; 9905-0137724;
PR 08-JUN-1999; 9905-0138094;
PR 10-JUN-1999; 9905-0138847;
PR 14-JUN-1999; 9905-0139119;
PR 16-JUN-1999; 9905-0139452;
PR 16-JUN-1999; 9905-0139453;
PR 17-JUN-1999; 9905-0139492;
PR 18-JUN-1999; 9905-0139454;
PR 18-JUN-1999; 9905-0139455;
PR 18-JUN-1999; 9905-0139456;
PR 18-JUN-1999; 9905-0139457;
PR 18-JUN-1999; 9905-0139458;
PR 18-JUN-1999; 9905-0139459;
PR 18-JUN-1999; 9905-0139460;
PR 18-JUN-1999; 9905-0139461;
PR 18-JUN-1999; 9905-0139462;
PR 18-JUN-1999; 9905-0139463;
PR 18-JUN-1999; 9905-0139750;
PR 18-JUN-1999; 9905-0139763;
PR 21-JUN-1999; 9905-0139817;
PR 22-JUN-1999; 9905-0139899;
PR 23-JUN-1999; 9905-0140353;
PR 23-JUN-1999; 9905-0140354;
PR 24-JUN-1999; 9905-0140695;
PR 28-JUN-1999; 9905-0140923;
PR 29-JUN-1999; 9905-0140991;
PR 30-JUN-1999; 9905-0141287;
PR 01-JUL-1999; 9905-0141942;
PR 01-JUL-1999; 9905-0142154;
PR 02-JUL-1999; 9905-0142055;
PR 06-JUL-1999; 9905-0142380;
PR 08-JUL-1999; 9905-0142803;
PR 09-JUL-1999; 9905-0142920;
PR 12-JUL-1999; 9905-0142977;
PR 13-JUL-1999; 9905-0143542;
PR 14-JUL-1999; 9905-0143624;

[illegible]

PP	06-OCT-1999;	99US-0157865.
PP	07-OCT-1999;	99US-0158029.
PP	08-OCT-1999;	99US-0158232.
PP	12-OCT-1999;	99US-0158369.
PP	13-OCT-1999;	99US-0159293.
PP	13-OCT-1999;	99US-0159294.
PP	13-OCT-1999;	99US-0159295.
PP	14-OCT-1999;	99US-0159329.
PP	14-OCT-1999;	99US-0159330.
PP	14-OCT-1999;	99US-0159331.
PP	14-OCT-1999;	99US-0159637.
PP	14-OCT-1999;	99US-0159639.
PP	19-OCT-1999;	99US-0159584.
PP	21-OCT-1999;	99US-0160741.
PP	21-OCT-1999;	99US-0160767.
PP	21-OCT-1999;	99US-0160778.
PP	21-OCT-1999;	99US-0160814.
PP	21-OCT-1999;	99US-0160815.
PP	22-OCT-1999;	99US-0160980.
PP	22-OCT-1999;	99US-0160981.
PP	27-OCT-1999;	99US-0160989.
PP	25-OCT-1999;	99US-0161404.
PP	25-OCT-1999;	99US-0161405.
PP	25-OCT-1999;	99US-0161406.
PP	26-OCT-1999;	99US-0161359.
PP	26-OCT-1999;	99US-0161360.
PP	26-OCT-1999;	99US-0161361.
PP	28-OCT-1999;	99US-0161920.
PP	28-OCT-1999;	99US-0161922.
PP	28-OCT-1999;	99US-0161923.
PP	29-OCT-1999;	99US-0162142.

Query Match	91.18	Score 41	EB 21	Length 325
Best Local Similarity	83.34	Pred. No. 45		
Matches	5	Conservative	0	Mismatches 1
				Indels 0
QY	1	WVRWHF	6	
Db	141	WVRWHF	146	

```

RESULT 8
AMG48292
ID    AMG48292 standard; Protein, 205 AA.
XX
AC    AMG48292;
XX
RT    18 OCT-2000 (first entry)
XX
DE    Arabidopsis thaliana protein fragment SEQ ID NO: 60968.
XX
KW    Protein identification; signal transduction pathway; metabolic pathway;
KW    hybridisation assay; genetic mapping; gene expression control; promoter;
KW    termination sequence.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405 A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000; 2000EP-0301439.
XX
PR    25 FEB 1999; 99US-0121825.
PR    05-MAR-1999; 99US-0123180.
PR    09 MAR 1999; 99US-0123548.
PR    23 MAR 1999; 99US-0125788.
PR    25 MAR 1999; 99US-0126264.
PR    29-MAR-1999; 99US-0126785.
PR    01 APR 1999; 99US-0127462.
PR    06 APR 1999; 99US-0128234.
PR    08 APR 1999; 99US-0128714.

```

[illegible]

PR 14-MAY-1999; 99US-0134218
PR 14-MAY-1999; 99US-0134219
PR 14-MAY-1999; 99US-0134221
PR 14-MAY-1999; 99US-0134370
PR 18-MAY-1999; 99US-0134768
PR 19-MAY-1999; 99US-0134941
PR 20-MAY-1999; 99US-0135124
PR 21-MAY-1999; 99US-0135353
PR 24-MAY-1999; 99US-0135529
PR 25-MAY-1999; 99US-0136021
PR 27-MAY-1999; 99US-0136392
PR 28-MAY-1999; 99US-0136782
PR 01-JUN-1999; 99US-0137022
PR 03-JUN-1999; 99US-0137528
PR 04-JUN-1999; 99US-0137562
PR 07-JUN-1999; 99US-0137724
PR 08-JUN-1999; 99US-0138094
PR 10-JUN-1999; 99US-0138540
PR 10-JUN-1999; 99US-0138847
PR 14-JUN-1999; 99US-0139119
PR 16-JUN-1999; 99US-0139452
PR 16-JUN-1999; 99US-0139453
PR 17-JUN-1999; 99US-0139492
PR 18-JUN-1999; 99US-0139454
PR 18-JUN-1999; 99US-0139455
PR 18-JUN-1999; 99US-0139456
PR 18-JUN-1999; 99US-0139457
PR 18-JUN-1999; 99US-0139458
PR 18-JUN-1999; 99US-0139459
PR 18-JUN-1999; 99US-0139460
PR 18-JUN-1999; 99US-0139461
PR 18-JUN-1999; 99US-0139462
PR 18-JUN-1999; 99US-0139463
PR 18-JUN-1999; 99US-0139760
PR 18-JUN-1999; 99US-0139763
PR 21-JUN-1999; 99US-0139817
PR 22-JUN-1999; 99US-0139899
PR 23-JUN-1999; 99US-0140353
PR 23-JUN-1999; 99US-0140354
PR 24-JUN-1999; 99US-0140646
PR 28-JUN-1999; 99US-0140823
PR 29-JUN-1999; 99US-0140991
PR 30-JUN-1999; 99US-0141287
PR 01-JUL-1999; 99US-0141842
PR 01-JUL-1999; 99US-0142154
PR 02-JUL-1999; 99US-0142056
PR 06-JUL-1999; 99US-0142390
PR 08-JUL-1999; 99US-0142803
PR 09-JUL-1999; 99US-0142920
PR 12-JUL-1999; 99US-0142977
PR 13-JUL-1999; 99US-0143542
PR 14-JUL-1999; 99US-0143624
PR 15-JUL-1999; 99US-0144005
PR 16-JUL-1999; 99US-0144085
PR 16-JUL-1999; 99US-0144086
PR 19-JUL-1999; 99US-0144325
PR 19-JUL-1999; 99US-0144331
PR 19-JUL-1999; 99US-0144332
PR 19-JUL-1999; 99US-0144333
PR 19-JUL-1999; 99US-0144334
PR 19-JUL-1999; 99US-0144335
PR 20-JUL-1999; 99US-0144352
PR 20-JUL-1999; 99US-0144632
PR 20-JUL-1999; 99US-0144884
PR 21-JUL-1999; 99US-0144814
PR 21-JUL-1999; 99US-0145086
PR 21-JUL-1999; 99US-0145088
PR 22-JUL-1999; 99US-0145085
PR 22-JUL-1999; 99US-0145087
PR 22-JUL-1999; 99US-0145089
PR 22-JUL-1999; 99US-0145102
PR 23-JUL-1999; 99US-0145145
PR 23-JUL-1999; 99US-0145218

PR 23-JUL-1999; 99US-0145224
PR 26-JUL-1999; 99US-0145276
PR 27-JUL-1999; 99US-0145913
PR 27-JUL-1999; 99US-0145918
PR 27-JUL-1999; 99US-0145919
PR 28-JUL-1999; 99US-0145951
PR 02-AUG-1999; 99US-0146386
PR 07-AUG-1999; 99US-0146388
PR 07-AUG-1999; 99US-0146389
PR 03-AUG-1999; 99US-0147038
PR 04-AUG-1999; 99US-0147204
PR 04-AUG-1999; 99US-0147302
PR 05-AUG-1999; 99US-0147302
PR 05-AUG-1999; 99US-0147260
PR 06-AUG-1999; 99US-0147303
PR 06-AUG-1999; 99US-0147416
PR 09-AUG-1999; 99US-0147493
PR 09-AUG-1999; 99US-0147935
PR 10-AUG-1999; 99US-0148171
PR 11-AUG-1999; 99US-0148319
PR 12-AUG-1999; 99US-0148341
PR 13-AUG-1999; 99US-0148565
PR 13-AUG-1999; 99US-0148684
PR 16-AUG-1999; 99US-0149368
PR 17-AUG-1999; 99US-0149175
PR 18-AUG-1999; 99US-0149426
PR 20-AUG-1999; 99US-0149722
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149929
PR 23-AUG-1999; 99US-0149992
PR 23-AUG-1999; 99US-0149993
PR 23-AUG-1999; 99US-0149994
PR 25-AUG-1999; 99US-0150566
PR 26-AUG-1999; 99US-0150884
PR 27-AUG-1999; 99US-0151065
PR 27-AUG-1999; 99US-0151066
PR 27-AUG-1999; 99US-0151080
PR 30-AUG-1999; 99US-0151303
PR 31-AUG-1999; 99US-0151438
PR 01-SEP-1999; 99US-0151930
PR 07-SEP-1999; 99US-0152363
PR 10-SEP-1999; 99US-0153070
PR 13-SEP-1999; 99US-0153758
PR 15-SEP-1999; 99US-0154018
PR 16-SEP-1999; 99US-0154039
PR 20-SEP-1999; 99US-0154779
PR 22-SEP-1999; 99US-0155139
PR 23-SEP-1999; 99US-0155486
PR 24-SEP-1999; 99US-0155659
PR 28-SEP-1999; 99US-0156458
PR 29-SEP-1999; 99US-0156596
PR 04-OCT-1999; 99US-0157117
PR 05-OCT-1999; 99US-0157753
PR 06-OCT-1999; 99US-0157865
PR 07-OCT-1999; 99US-0158029
PR 08-OCT-1999; 99US-0158232
PR 12-OCT-1999; 99US-0158369
PR 13-OCT-1999; 99US-0159293
PR 13-OCT-1999; 99US-0159294
PR 13-OCT-1999; 99US-0159295
PR 14-OCT-1999; 99US-0159329
PR 14-OCT-1999; 99US-0159330
PR 14-OCT-1999; 99US-0159331
PR 14-OCT-1999; 99US-0159637
PR 14-OCT-1999; 99US-0159638
PR 18-OCT-1999; 99US-0159584
PR 21-OCT-1999; 99US-0160741
PR 21-OCT-1999; 99US-0160767
PR 21-OCT-1999; 99US-0160768
PR 21-OCT-1999; 99US-0160770
PR 21-OCT-1999; 99US-0160814
PR 21-OCT-1999; 99US-0160815
PR 22-OCT-1999; 99US-0160980
PR 22-OCT-1999; 99US-0160981

PR	01-OCT-1999;	99US-0147004
PR	02-AUG-1999;	99US-0146380
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-0147204
PR	04-AUG-1999;	99US-0147302
PR	05-AUG-1999;	99US-0147143
PR	05-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	06-AUG-1999;	99US-0147416
PR	09-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148317
PR	11-AUG-1999;	99US-0148171
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149168
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149728
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	25-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150804
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	07-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154739
PR	22-SEP-1999;	99US-0155132
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156459
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0158392
PR	13-OCT-1999;	99US-0159294
PR	13-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159684
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404
PR	25-OCT-1999;	99US-0161405
PR	25-OCT-1999;	99US-0161406
PR	26-OCT-1999;	99US-0161359
PR	26-OCT-1999;	99US-0161360
PR	26-OCT-1999;	99US-0161361

PP 28-OCT-1999; 99US-0161920;
PP 28-OCT-1999; 99US-0161999;
PP 28-OCT-1999; 99US-0161999;
PP 29-OCT-1999; 99US-0161942;

Query Match	98.38%	Score 407	SP 217	Length 338
Best Local Similarity	93.38%	Prod No 67		
Matches	5	Conservative	0	Mismatches 1
				Indels 0
				Gaps 0

QY	1	WVRWHF	6
Db	147	WVNVHF	152

RESULT	12
AAG47193	
ID	AAG47193 standard, Protein, 338 AA
XX	
AC	AAG47193;
XX	
DT	18 OCT 2000 (first entry)

DE Arabidopsis thaliana protein fragment SPQ TP No. 59454.

Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

Arabidopsis thaliana.

PN EPI 033405-A2.

PD 06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

PR	25-FEB-1999;	99US-0121825
PR	05-MAR-1999;	99US-0123180
PR	09-MAR-1999;	99US-0123548
PR	23-MAR-1999;	99US-0125788
PR	25-MAR-1999;	99US-0126264
PR	29-MAR-1999;	99US-0126785
PR	01-APR-1999;	99US-0127462
PR	08-APR-1999;	99US-0128234
PR	16-APR-1999;	99US-0128714
PR	19-APR-1999;	99US-0130077
PR	21-APR-1999;	99US-0130549
PR	23-APR-1999;	99US-0130510
PR	23-APR-1999;	99US-0130891
PR	28-APR-1999;	99US-0131449
PR	30-APR-1999;	99US-0132048
PR	30-APR-1999;	99US-0132407
PR	04-MAY-1999;	99US-0132484
PR	05-MAY-1999;	99US-0132485
PR	06-MAY-1999;	99US-0132486
PR	07-MAY-1999;	99US-0132863
PR	11-MAY-1999;	99US-0132864
PR	14-MAY-1999;	99US-0134218
PR	14-MAY-1999;	99US-0134219
PR	14-MAY-1999;	99US-0134221
PR	14-MAY-1999;	99US-0134370
PR	18-MAY-1999;	99US-0134378
PR	19-MAY-1999;	99US-0134941
PR	20-MAY-1999;	99US-0135124
PR	21-MAY-1999;	99US-0135353
PR	24-MAY-1999;	99US-0135629
PR	25-MAY-1999;	99US-0136021
PR	27-MAY-1999;	99US-0136392
PR	28-MAY-1999;	99US-0136782
PR	01-JUN-1999;	99US-0137293
PR	03-JUN-1999;	99US-0137528

Matches 5; Conservatve 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 WVRWHF 6
 Db 147 WVNWHF 152

RESULT 13

AAR60501
 ID AAR60501 standard; Protein; 418 AA.

XX AAR60501;

XX 2a-MAP-1994 (first entry)

XX linoleic-acid-desaturase

XX linoleic-acid-desaturase; oleic-desaturase; transgenic plant;

XX crop improvement; linolenic acid.

XX Not specified.

XX WO9418337-A.

XX 18-AUG-1994.

XX 04-FEB-1994; 44WO-0501321

XX 05-FEB-1994; 93US-0014431

XX 22 NOV 1993; 93US 0156551

XX (MONS) MONSANTO CO

XX (UNMS) UNIV MICHIGAN STATE

XX Arcotel VIA, Gibson St, Kishore GM, Puff TG, Somerville CR;

XX WPI; 1994-279758/34.

XX N FSDDB; AAQ71242.

XX Genetically transformed plants with altered linolenic acid

XX content - plant recombinant; double stranded RNA encoding

XX linoleic acid desaturase, or the antisense of the coding

XX sequence 418 AA;

XX Disclosure: Page 100-102; 144pp; English

XX The isolation of cDNA encoding linoleic-acid-desaturase and

XX oleic-desaturase is described. A sequence of the invention is

XX given in AAQ71242, and its encoded protein sequence in AAR60501

XX Sequence 418 AA;

RESULT 14

AAQ05808
 ID AAQ05808 standard; Protein; 448 AA.

XX AAQ05808;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 7357

XX Protein identification: signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; growth;

KW termination sequence.

XX Arabidopsis thaliana

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135729.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137702.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 16-JUN-1999; 99US-0139453.

XX 17-JUN-1999; 99US-0139492.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139457.

XX 18-JUN-1999; 99US-0139458.

XX 18-JUN-1999; 99US-0139459.

XX 18-JUN-1999; 99US-0139460.

XX 18-JUN-1999; 99US-0139461.

XX 18-JUN-1999; 99US-0139462.

XX 18-JUN-1999; 99US-0139463.

XX 18-JUN-1999; 99US-0139750.

XX 18-JUN-1999; 99US-0139763.

XX 21-JUN-1999; 99US-0139817.

XX 22-JUN-1999; 99US-0139899.

XX 23-JUN-1999; 99US-0140353.

XX 23-JUN-1999; 99US-0140354.

XX 24-JUN-1999; 99US-0140695.

XX
 PR 25-FEB-1999; 99US-0121925
 XX
 PR 05-MAR-1999; 99US-0123180
 PR 09-MAR-1999; 99US-0123548
 PR 23-MAR-1999; 99US-0125788
 PR 25-MAR-1999; 99US-0126264
 PR 29-MAR-1999; 99US-0126785
 PR 01-APR-1999; 99US-0127462
 PR 06-APR-1999; 99US-0128234
 PR 08-APR-1999; 99US-0128714
 PR 16-APR-1999; 99US-0129845
 PR 19-APR-1999; 99US-0130077
 PR 21-APR-1999; 99US-0130449
 PR 23-APR-1999; 99US-0130610
 PR 23-APR-1999; 99US-0130891
 PR 28-APR-1999; 99US-0131449
 PR 30-APR-1999; 99US-0132048
 PR 04-MAY-1999; 99US-0132407
 PR 05-MAY-1999; 99US-0132484
 PR 06-MAY-1999; 99US-0132485
 PR 06-MAY-1999; 99US-0132486
 PR 06-MAY-1999; 99US-0132487
 PR 07-MAY-1999; 99US-0132863
 PR 11-MAY-1999; 99US-0134256
 PR 14-MAY-1999; 99US-0134218
 PR 14-MAY-1999; 99US-0134219
 PR 14-MAY-1999; 99US-0134221
 PR 14-MAY-1999; 99US-0134370
 PR 18-MAY-1999; 99US-0134768
 PR 19-MAY-1999; 99US-0134941
 PR 20-MAY-1999; 99US-0135124
 PR 21-MAY-1999; 99US-0135363
 PR 24-MAY-1999; 99US-0135629
 PR 25-MAY-1999; 99US-0136021
 PR 27-MAY-1999; 99US-0136392
 PR 28-MAY-1999; 99US-0136792
 PR 01-JUN-1999; 99US-0137222
 PR 03-JUN-1999; 99US-0137528
 PR 04-JUN-1999; 99US-0137502
 PR 07-JUN-1999; 99US-0137724
 PR 08-JUN-1999; 99US-0138034
 PR 10-JUN-1999; 99US-0138540
 PR 10-JUN-1999; 99US-0138847
 PR 14-JUN-1999; 99US-0139119
 PR 16-JUN-1999; 99US-0139452
 PR 16-JUN-1999; 99US-0139453
 PR 17-JUN-1999; 99US-0139492
 PR 18-JUN-1999; 99US-0139454
 PR 18-JUN-1999; 99US-0139455
 PR 18-JUN-1999; 99US-0139456
 PR 18-JUN-1999; 99US-0139457
 PR 18-JUN-1999; 99US-0139458
 PR 18-JUN-1999; 99US-0139459
 PR 18-JUN-1999; 99US-0139460
 PR 18-JUN-1999; 99US-0139461
 PR 18-JUN-1999; 99US-0139462
 PR 18-JUN-1999; 99US-0139463
 PR 18-JUN-1999; 99US-0139750
 PR 18-JUN-1999; 99US-0139763
 PR 21-JUN-1999; 99US-0139817
 PR 22-JUN-1999; 99US-0139899
 PR 23-JUN-1999; 99US-0140353
 PR 23-JUN-1999; 99US-0140354
 PR 24-JUN-1999; 99US-0140695
 PR 28-JUN-1999; 99US-0140823
 PR 29-JUN-1999; 99US-0140991
 PR 30-JUN-1999; 99US-0141287
 PR 01-JUL-1999; 99US-0141842
 PR 01-JUL-1999; 99US-0142154
 PR 02-JUL-1999; 99US-0142066
 PR 05-JUL-1999; 99US-0142390

PR 08-JUL-1999; 99US-0142803
 PR 09-JUL-1999; 99US-0142920
 PR 12-JUL-1999; 99US-0142977
 PR 13-JUL-1999; 99US-0143542
 PR 14-JUL-1999; 99US-0143624
 PR 15-JUL-1999; 99US-0144005
 PR 16-JUL-1999; 99US-0144085
 PR 16-JUL-1999; 99US-0144086
 PR 19-JUL-1999; 99US-0144325
 PR 19-JUL-1999; 99US-0144331
 PR 19-JUL-1999; 99US-0144332
 PR 19-JUL-1999; 99US-0144333
 PR 19-JUL-1999; 99US-0144334
 PR 19-JUL-1999; 99US-0144335
 PR 20-JUL-1999; 99US-0144352
 PR 20-JUL-1999; 99US-0144632
 PR 20-JUL-1999; 99US-0144884
 PR 21-JUL-1999; 99US-0144814
 PR 21-JUL-1999; 99US-0145086
 PR 21-JUL-1999; 99US-0145088
 PR 22-JUL-1999; 99US-0145087
 PR 22-JUL-1999; 99US-0145089
 PR 22-JUL-1999; 99US-0145192
 PR 23-JUL-1999; 99US-0145145
 PR 23-JUL-1999; 99US-0145218
 PR 23-JUL-1999; 99US-0145218
 PR 26-JUL-1999; 99US-0145224
 PR 26-JUL-1999; 99US-0145276
 PR 27-JUL-1999; 99US-0145913
 PR 27-JUL-1999; 99US-0145918
 PR 27-JUL-1999; 99US-0145919
 PR 28-JUL-1999; 99US-0145951
 PR 02-AUG-1999; 99US-0146386
 PR 02-AUG-1999; 99US-0146388
 PR 02-AUG-1999; 99US-0146389
 PR 03-AUG-1999; 99US-0147038
 PR 04-AUG-1999; 99US-0147204
 PR 04-AUG-1999; 99US-0147302
 PR 05-AUG-1999; 99US-0147192
 PR 05-AUG-1999; 99US-0147260
 PR 06-AUG-1999; 99US-0147303
 PR 06-AUG-1999; 99US-0147416
 PR 09-AUG-1999; 99US-0147493
 PR 09-AUG-1999; 99US-0147935
 PR 10-AUG-1999; 99US-0148171
 PR 11-AUG-1999; 99US-0148319
 PR 12-AUG-1999; 99US-0148341
 PR 13-AUG-1999; 99US-0148565
 PR 13-AUG-1999; 99US-0148684
 PR 16-AUG-1999; 99US-0149368
 PR 17-AUG-1999; 99US-0149175
 PR 18-AUG-1999; 99US-0149426
 PR 20-AUG-1999; 99US-0149722
 PR 20-AUG-1999; 99US-0149723
 PR 20-AUG-1999; 99US-0149929
 PR 21-AUG-1999; 99US-0149962
 PR 23-AUG-1999; 99US-0149930
 PR 25-AUG-1999; 99US-0150566
 PR 26-AUG-1999; 99US-0150884
 PR 27-AUG-1999; 99US-0151065
 PR 27-AUG-1999; 99US-0151066
 PR 27-AUG-1999; 99US-0151080
 PR 30-AUG-1999; 99US-0151303
 PR 31-AUG-1999; 99US-0151438
 PR 01-SEP-1999; 99US-0151930
 PR 07-SEP-1999; 99US-0152363
 PR 10-SEP-1999; 99US-0153070
 PR 13-SEP-1999; 99US-0153758
 PR 15-SEP-1999; 99US-0154018
 PR 16-SEP-1999; 99US-0154039
 PR 20-SEP-1999; 99US-0154779
 PR 22-SEP-1999; 99US-0155139
 PR 23-SEP-1999; 99US-0155486

PR 04-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157153.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 10-OCT-1999; 99US-0158369.
 PR 11-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160778.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 88.9%; Score 40; DB 21; Length 448;
 Recd: 2003; Similarity: 93.3%; Pred. NO. 90;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVRWHF 6
 DB 257 WVRWHF 262

Search completed: February 20, 2003, 10:47:34
 Tot time: 30.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:41:10, Search time 12 seconds
(without alignments)
48,067 million cell updates/sec

Title: US-09-912-414-2
Perfect score: 45
Sequence: 1 WVPWHF 6

Scoring table: PROSUMMARY
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIP 71.1
1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Pred. No. is the number of results potentially changed by having a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	88.9	443	2 T08136	probable omega-6
2	40	88.9	448	2 Pae362	hypothetical prote
3	39	86.7	291	2 A83299	hypothetical prote
4	39	86.7	370	2 A96741	hypothetical prote
5	39	86.7	492	2 S49466	catalase (EC 1.11.
6	38	84.4	477	1 I73631	brain-derived neur
7	38	84.4	822	1 A56853	brain-derived neur
8	37	82.2	491	2 C01586	probable protein 5
9	36	80.0	185	2 T44538	hypothetical prote
10	36	80.0	230	2 D70847	hypothetical prote
11	36	80.0	262	2 H83214	probable peroxide
12	36	80.0	265	2 A00441	probable aliphatic
13	36	80.0	266	2 AD3239	hypothetical prote
14	36	80.0	294	2 E75295	conserved hypothet
15	36	80.0	424	2 T07742	omega-6 desaturase
16	36	80.0	472	2 AG2417	hypothetical prote
17	36	80.0	754	2 AF0614	hypothetical prote
18	36	80.0	1041	2 T29010	hypothetical prote
19	35	77.8	111	2 AB0690	probable membrane
20	35	77.8	226	1 F64450	hypothetical prote
21	35	77.8	458	2 B81409	probable transmemb
22	35	77.8	509	2 C00372	4-hydroxyphenylase
23	35	77.8	718	2 T29448	hypothetical prote
24	34	75.6	96	4 Q0EC31	hypothetical prote
25	34	75.6	151	2 E64430	ribosomal protein
26	34	75.6	179	2 P81344	hypothetical prote
27	34	75.6	219	2 T27954	hypothetical prote
28	34	75.6	265	2 A99432	conserved hypothet
29	34	75.6	301	2 G83352	transcription regu

30	34	75.6	316	2 C47029	155F hemolysin, nat
31	34	75.6	316	2 C64963	transcription requ
32	34	75.6	316	2 E85823	transcription requ
33	34	75.6	316	2 G90976	transcription requ
34	34	75.6	329	2 A81891	hypothetical prote
35	34	75.6	358	1 S43876	site-specific DNA
36	34	75.6	358	2 A97036	modification methyl
37	34	75.6	424	2 D40055	partial E. coli DNA-h
38	34	75.6	424	2 J05891	omega-6 desaturase
39	34	75.6	495	2 C05891	active esterase (fla
40	34	75.6	497	2 T40586	nucleolar protein
41	34	75.6	502	2 T40586	nucleolar protein
42	34	75.6	509	2 G81104	hypothetical prote
43	34	75.6	709	2 T48889	serine/threonine p
44	34	75.6	910	2 C69456	subtilisin sendai
45	33.5	74.4	1511	2 C69032	probable membrane

ALIGNMENTS

RESULT 1

T08136
probable omega-6 desaturase (FC 1 14 99 -) precursor, chloroplast - rape

C/Species: Brassica napus (rape)

C/Date: 21 May 1999 #sequence_revision 21-May-1999 #text_change 08 Oct 1999

C/Accession: T08136

F.Hitz, W.D., Carlson, T.S., Borth, J.P., Kinney, A.C., Stewart, P.B., Taylor, N.S.
Plant Physiol. 105, 635-641, 1994

A/Title: Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and the oxy

A/Reference number: Z16109; MUID:94345008; PMID:8066133

A/Accession: T08136

A/Status: preliminary, translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-443 <HIT>

A/Cross-references: EMBL:Z16109, MUID:94345008, PMID:8066133

A/Experimental source: seed

C/Genetics:

A/Genome: nuclear

C/Superfamily: omega-3 fatty acid desaturase

C/Keywords: chloroplast; fatty acid metabolism; oxidoreductase

Query Match	88.9%	Score 40	DB 2	Length 443
Best Local Similarity	83.3%	Pred. No. 31		
Matches	5	Conservative	0	Mismatches 1
DB	252 WVPWHF	257		

RESULT 2

D95362
hypothetical protein AT4930950 (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse ear cress)

C/Date: 16 Feb 2001 #sequence_revision 16 Feb 2001 #text_change 17 May-2002

C/Accession: D95362

F.Anczyk, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A/Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A/Reference number: A85001, MUID:9292498, PMID:10617198

A/Accession: D95362

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-448 <STO>

A/Cross-references: GB:NC_001078, MUID:9707477, FID: C47141.1, MUID: 9707477

C/Genetics:

A/Gene: AT4930950

A/Map position: 4

C/Superfamily: omega-3 fatty acid desaturase

Query Match: 88.9%, Score 40, DB 2, Length 443;
Best Local Similarity: 83.3%; Pred. No. 31;

E453-477/Domain	cytosolic	#startus predicted <CYT>	E453-477/Domain	cytosolic	#startus experimental
F1,32-38,36-40,152	176,154	194	219	266,302	345/151
F2,39-45,41-44,164	264,260,225	339,359	410/Binding site	carbohydrate (Asn)	glycolipid
F3,46-52,53-59,179	205,241,264,260,225	339,359	410/Binding site	carbohydrate (Asn)	glycolipid

Query Match	84.48;	Score 38;	DB 1;	Length 477;
Best Local Similarity	80.08;	Pred. No. 69;		
Matches	4;	Conservative	1;	Mismatches 0;
				Indels 0;
				Gaps 0;

QY	1	WVRWH	5
		1:111	
Dh	4	WIRWH	8

RESULT 7

A566853
brain-derived neurotrophic factor receptor precursor human
N/Alternate names: receptor tyrosine kinase trkB
N/Contains: protein-tyrosine kinase (EC 2.7.1.12)
C/Species: Homo sapiens (man)
C/Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 11-Jun-1999
C/Accession: A566853; 156557
P;Nagagawara, A.; Lin, X.G.; Ikegaki, N.; White, P.S.; Yamashita, D.J.; Nyeum, L.M.; Bie
Genomics 25, 538-546, 1995
A/Title: Cloning and chromosomal localization of the human TRK-B tyrosine kinase receptor
A/Reference number: A566853; MUID:95103022; PMID:7789398

```
Query Match      84.4%; Score 38; DB 1; Length 822;
Best Local Similarity 80.0%; Pred No 1 1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	WVRWH
		:
Db	4	WIRWH

RESULT 8

G01586
 Ensembl Protein ID: 1; Homology: 1 (similarity) - human
 C/Species: Homo sapiens (man)
 C/Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #ext_change 20-Jun-2000
 C/Accession: G01586

Query Match	82.2%	Score 37	DP 2	Length 591
Post Local similarity	66.7%	Prod W-1.0e+03		
Matches	4	Conservative	1	Mismatches 1
				Indels 0
				Gaps 0

QY	1	WVRWHF	6
		:	
Db	350	WCRWHY	355

RESULT 9
T44538
Myxobolus [a] protein VP2 [Imported] - Pseudomonas aeruginosa

Query Match	80.0%	Score 36;	DP 2;	Length 185;
Post Local Similarity	80.0%	Pred. No. 60;		
Matches	4;	Conservative	1;	Mismatch-Bias
			0;	Indels
			0;	Gaps
			0	

QY	1	WVRWH	5
		:	
Db	42	WLRWH	46

RESULT 10
D70847

hypothetical protein Rv0059 - Mycobacterium tuberculosis (strain H37RV)

C; Species: *Mycobacterium tuberculosis*

C;Date: 17-Jul-1998 #sequence 17-Jul-1998 #text_change 20-Jun-2000

C:\Accession: D70847

R. Cole, S. T. Brosch, R. Parkhill, J. Garnier, T. Churcher, C. Harris, D. Gordon, S. Connor, R. Davies, R. Devlin, K. Felwell, T. Gentles, S. Hamlin, N. Holroyd, S. Rajandream, M. A. Rogers, J. Rutter, S. Seeger, K. Skelton, S. Squares, S.

Nature 392, 577-514, 1999
A/Author: Sgares, R.; Salston, J. F.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Localizing the biology of *Yersinia enterocolitica* using the complete genome
A/Reference number: Nature, 392, 577-514, 1999
A/Accession: D9947
A/Status: Preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-230 <COL>
A/Accession: GB:AC00449, CP:AF00449, MF:J00449, F00449, F00449, F00449
A/Experimental source: strain H27PV
C/Genetics:
A/Genes: RY0059
C/Superfamily: Synchococcus nitrate transport protein ntrB

Query Match: 60.0%; Score 36; DB 2; Length 252;
Best Local Similarity: 60.0%; Pred. No. 82;
Matches: 4; Conserved: 1; Mismatches: 0; Indels: 0; Gaps: 0

QY 1 WVRWH 5
|:|:|
DB 248 WVRWH 252

RESULT 11
AC0441
A/Title: Complete genome sequence of *Yersinia enterocolitica* strain H27PV, an opportunistic pathogen
A/Reference number: Nature, 392, 577-514, 1999
A/Accession: H83214
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-262 <COL>
A/Cross-references: GB:AE004765; GB:AE004091; NID:G9449580; PID:AA006831.1; GSDB:GN001
A/Experimental source: strain H27PV
C/Genetics:
A/Genes: PA3443
C/Superfamily: Synchococcus nitrate transport protein ntrB

Query Match: 60.0%; Score 36; DB 2; Length 252;
Best Local Similarity: 60.0%; Pred. No. 82;
Matches: 4; Conserved: 1; Mismatches: 0; Indels: 0; Gaps: 0

QY 1 WVRWH 5
|:|:|
DB 248 WVRWH 252

RESULT 12
AC0441
A/Title: Complete genome sequence of *Yersinia enterocolitica* strain H27PV, an opportunistic pathogen
A/Reference number: Nature, 392, 577-514, 1999
A/Accession: AC0441
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-262 <COL>
A/Cross-references: GB:AE004765; GB:AE004091; NID:G9449580; PID:AA006831.1; GSDB:GN001
A/Experimental source: strain H27PV
C/Genetics:
A/Genes: PA3443
C/Superfamily: Synchococcus nitrate transport protein ntrB

C/Superfamily: Synchococcus nitrate transport protein ntrB

Query Match: 60.0%; Score 36; DB 2; Length 252;
Best Local Similarity: 60.0%; Pred. No. 82;
Matches: 4; Conserved: 1; Mismatches: 0; Indels: 0; Gaps: 0

QY 1 WVRWH 5
|:|:|
DB 248 WVRWH 252

RESULT 13
AD3238
A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A/Reference number: Nature, 392, 577-514, 1999
A/Accession: AD3238
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-262 <COL>
A/Cross-references: GB:AE004765; GB:AE004091; NID:G9449580; PID:AA006831.1; GSDB:GN001
A/Experimental source: strain C58 (Colombia)
C/Genetics:
A/Genes: ntrB
C/Superfamily: Synchococcus nitrate transport protein ntrB

Query Match: 60.0%; Score 36; DB 2; Length 252;
Best Local Similarity: 60.0%; Pred. No. 82;
Matches: 4; Conserved: 1; Mismatches: 0; Indels: 0; Gaps: 0

QY 1 WVRWH 5
|:|:|
DB 257 WVRWH 261

RESULT 14
E70295
A/Title: Complete genome sequence of the radiodurant bacterium *Deinococcus radiodurans* R1.
A/Reference number: Nature, 392, 577-514, 1999
A/Accession: E70295
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-262 <COL>
A/Cross-references: GB:AE004765; GB:AE004091; NID:G9449580; PID:AA006831.1; GSDB:GN001
A/Experimental source: strain R1
C/Genetics:
A/Genes: PF0253
A/Map position: 1

Query Match: 60.0%; Score 36; DB 2; Length 384;
Best Local Similarity: 60.0%; Pred. No. 102;
Matches: 4; Conserved: 1; Mismatches: 0; Indels: 0; Gaps: 0

QY 1 WVRWH 5
|:|:|

DB 362 WLRWH 366

RESULT 15

T07742

omega-6 desaturase, chloroplast - soybean

C:Species: Glycine max (soybean)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999

C:Accession: T07742

R:Hiltz, W.D.; Carlson, T.J.; Roach, J.P.; Kinney, A.J.; Stecca, R.L.; Yadav, N.S.

Plant Physiol. 105, 635-641, 1994

A:Title: Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its ex

A:Reference number: 216103; MIM:34345008; PMID:8066133

A:Accession: T07742

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-424 <HIT>

A:Cross-references: EMBL:L29215; NID:G459061, PIRN AAA50158.1, PIR:G459962

A:Experimental source: seed

C:Genetics:

A:Genome: nuclear

C:Superfamily: omega-3 fatty acid desaturase

C:Keywords: chloroplast

Query Match 80.0%; Score 36; DB 2; Length 424;
Best local similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVRWHF 6

DB 251 WLRWHF 256

Search completed: February 20, 2003, 10:50:00
Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1991 - 2003 Copyright Ltd

AM Protein - protein search, using SW model

Run on: February 20, 2003, 10:41:55, Search time 6.66667 seconds
(without alignment)
37 329 Million cell updates/sec

Title: US-09-912 414 2
Perfect score: 45
Sequence: 1 WVRWHF 6

Scoring table: BLOSUM62DX
Gapop 10 0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match length DB	ID	Description
1	42	93.3	479 1	CATA_PSEPU
2	40	88.9	443 1	EPAC_BPAWA
3	40	88.9	448 1	EPAC_ARATH
4	39	86.7	482 1	CATA_ONIVE
5	38	84.4	822 1	TPPR_HUMAN
6	36	80.0	424 1	FNAC_SOYBN
7	36	80.0	789 1	BGSS_XANAC
8	36	80.0	1169 1	SUV3_DROME
9	35	77.8	226 1	YC07_METJA
10	35	77.8	718 1	PLSR_CAEFL
11	34	75.6	148 1	PE12_METJA
12	34	75.6	316 1	CRU_ECOLI
13	34	75.6	316 1	CRU_KLEAF
14	34	75.6	358 1	MTG1_CAUOP
15	34	75.6	372 1	GP45_HUMAN
16	34	75.6	373 1	GP45_MOUSE
17	34	75.6	495 1	AGFN_MOUSE
18	34	75.6	799 1	AFSK_STPCC
19	34	75.6	807 1	AFSK_STGCP
20	34	75.6	1698 1	Y076_HUMAN
21	33.5	74.4	1511 1	PDRG_YEAST
22	33	73.3	244 1	Y168_CAEFL
23	33	73.3	376 1	MTS1_PHIME
24	33	73.3	377 1	MTB1_BRIAB
25	33	73.3	394 1	POP_ECOLI
26	33	73.3	439 1	Y112_ARATH
27	33	73.3	479 1	CATA_LACSK
28	33	73.3	483 1	BCA_STRVL
29	33	73.3	494 1	CATA_MICLU
30	33	73.3	502 1	CATA_TOXGO
31	33	73.3	504 1	SIKI_YEAST
32	33	73.3	511 1	NOD5_YEAST
33	33	73.3	529 1	NOD5_HUMAN

34	33	73.3	534 1	NOD5 RAT	O99286 ratius norv
35	33	73.3	614 1	Y03F_YEAST	P53293 saccharomyc
36	33	73.3	821 1	TEPR_MOUSE	P15009 mus musculu
37	33	73.3	821 1	TPPR_PAT	Q64604 ratius norv
38	33	73.3	1015 1	FDNG_ECOLI	P24183 escherichia
39	33	73.3	1016 1	FDNG_ECOLI	P27174 escherichia
40	32	71.1	129 1	YHME_YEAST	P38857 saccharomyc
41	32	71.1	146 1	Y03C_YEAST	P53311 saccharomyc
42	32	71.1	287 1	YAHF_ECOLI	P77207 escherichia
43	32	71.1	386 1	PLA_DRECA	P49165 urechis cau
44	32	71.1	395 1	DLTB_BACSU	P39580 bacillus su
45	32	71.1	396 1	RLA4_XENLA	P03427 xenopus lae

ALIGNMENTS

RESULT 1	
CATA_PSEPU	
ID CATA_PSEPU	STANDARD; PRT; 479 AA.
AC 059714;	
DT 30-MAY-2000 (Rel. 39, Created)	
DT 30-MAY-2000 (Rel. 39, Last sequence update)	
DT 30 MAY 2000 (Rel. 39, Last annotation update)	
DE Catalase (EC 1.11.1.6).	
GN KATA OR CATA.	
OS Pseudomonas putida.	
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;	
OC Pseudomonas.	
ON NCBI_TaxID=303;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=Corvallis;	
PY MERLIN=09010001, Pubmed=090009;	
PA Kim Y C, Miller C D, Anderson A J.	
RT "Identification of adjacent genes encoding the major catalase and a	
RT bacterioferritin from the plant-beneficial bacterium Pseudomonas	
RT putida";	
RL Gene 199:219-224(1997).	
CC FUNCTION: ENZYME: HYDROGEN PEROXIDE IN WATER AND HYDROGEN PEROXIDE.	
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.	
CC CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.	
CC COFACTOR: HEME GROUP.	
CC ENZYME REGULATION: ACTIVATED BY PEROXIDE.	
CC SIMILARITY: BELONGS TO THE CATALASE FAMILY.	
CC This SWISS-PROT entry is copyright. It is provided through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL Centre for	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (see http://www.isb-sib.ch/announcements/	
CC or send an email to license@sib-sib.ch).	
CC	
DR EMBL; U63511; AAB88219.1; ..	
DR HSSP; F42321; CAE.	
DR InterPro; IPR002226; Catalase.	
DR Pfam; PF00199; catalase; 1.	
DR PRINTS; PR00067; CATALASE.	
DR ProDom; PD000510; Catalase; 1.	
DR PROSITE; PS00437; CATALASE_1; 1.	
DR PROSITE; PS00438; CATALASE_2; 1.	
FW Oxidoreductase; peroxidase; iron; heme; hydrogen peroxide.	
FT ACT_SITE 53	BY SIMILARITY.
FT ACT_SITE 126	BY SIMILARITY.
FT BINDING 336	PROXIMAL HEME LIGAND (BY SIMILARITY).
FT SEQUENCE 479 AA; 53381 MW; PFF3CRDE67778571 CRC64;	
QY	Query Match
QY	Best Local Similarity 93.3%; Score 42, DB 1, Length 479;
QY	Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;


```

CC 1- FUNCTION: THE PLANT MEMBRANE FATTY ACID DESATURASE INTERFERES
CC THE SECOND FOURTEEN IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC ACIDS. IMMEDIATE CONSEQUENCES OF PLANT MEMBRANES. IT IS THOUGHT
CC TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC ESTERIFIED TO Glycerolipids, Sphingolipids AND PHOSPHATIDYLGLYCEROL.
CC 1- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC 1- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PEERABLE)
CC 1- DEVELOPMENTAL STAGE: HIGHEST LEVELS FOUND IN EXPANDING LEAVES.
CC 1- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/ OR BE INVOLVED IN METAL ION BINDING.
CC 1- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch)
CC
CC EMBL: U09503; AAA92800.1; -
CC EMBL: A1022198; CAI18198.1; -
CC EMBL: A1161578; CAB79813.1; -
CC EMBL: AY045621; AAY72979.1; -
CC EMBL: AY058078; AAI24196.1; -
CC EMBL: AY058852; AAI24240.1; -
CC InterPro: IPR001225; FA_desaturase.
CC Pfam: PF00487; FA_desaturase.1
CC ProDom: PD01081; FA_desaturase; 1
CC OXidoreductase; Fatty acid biosynthesis; Chloroplast; Mesotreme;
CC Transf. peptide.
CC TRANSIT 1 63 CHLOROPLAST (BY SIMILARITY).
CC CHAIN 70 448 OMEGA-6 FATTY ACID DESATURASE.
CC DOMAIN 171 175 HISTIDINE BOX-1.
CC DOMAIN 207 211 HISTIDINE BOX-2.
CC DOMAIN 367 371 HISTIDINE BOX-3.
CC SEQUENCE 444 AA; 11035 MW; 63457.982 kDa; 28264,
SQ
Query Match 88.9%; Score 40; DB 1; Length 448;
Best Local Similarity 92.3%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 WVRWHF 6
Db 257 WVRWHF 262

```

```

CC 1 CATALYTIC ACTIVITY: 2 H2O2(12) -> O2(12) + 2 H2O(10).
CC 1- COFACTOR: HEME GROUP.
CC 1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch)
CC
CC EMBL: X82176; CA857666.1; -
CC EMBL: AF069070; AAC79431.1; -
CC HESP: P42321; 2CAE
CC InterPro: IPR002226; Catalase.
CC Pfam: PF00199; catalase; 1.
CC PRINTS: PR00067; CATALASE
CC ProDom: PD000519; Catalase; 1.
CC PROSITE: PS00437; CATALASE_1; 1.
CC PROSITE: PS00438; CATALASE_2; 1.
CC OXidoreductase; Peroxidase; Heme, Hydrogen peroxide.
CC ACT SITE 55 55 BY SIMILARITY.
CC ACT SITE 128 128 BY SIMILARITY.
CC BINDING 338 338 PROXIMAL HEME LIGAND (BY SIMILARITY).
CC SEQUENCE 482 AA; 53700 MW; BIDA5E43BA39F776 CRC64;
SQ
Query Match 86.7%; Score 39; DB 1; Length 482;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WVRWHF 6
Db 211 WVRWHF 216

```

```

RESULT 4
CATA_ONCVE STANDARD, PRT, 482 AA.
AC Q27710; Q85499;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-MAY-2000 (Rel. 39, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN CAT.
OS Onchocerca volvulus (nematode).
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Wolbachiaceae
OC NCBI_TaxID=77551;
OX [1]
RN SEQUENCE FROM N.A.
RP Henkle-Duehrsen R.J.; Fekete V.H.G.; Wildenburg G.; Blaxter M.;
RA Walter R.D.;
RT "Gene structure, activity and localization of a catalase from
RT intracellular bacteria in Onchocerca volvulus";
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CT FUNCTION: HEME-DEPENDENT HYDROGEN PEROXIDE IN WATER AND CYTOSOL, SERVES
CT TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.

```

```

RESULT 5
TRKB_HUMAN STANDARD, PRT, 922 AA.
AC Q16620; Q16675;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE BDNF/NT-3 growth factors receptor precursor (PQ 2 7 1 112) (TrkB
DE tyrosine kinase) (GPI45-TrkB) (Trk-B).
GN NTRK2 OP TRKB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A. (LONG FORM).
RP Tissue=Hippocampus;
EX MEDLINE=95123473; PubMed=7789988;
PA Nakagawara A.; Liu X.-G.; Ikegaki N.; White P.S.; Yamashiro D.J.;
PA Nycum L.M.; Biegel J.A.; Brodeur G.M.;
RT "Cloning and chromosomal localization of the human Trk B tyrosine
RT kinase receptor gene (NTRK2).";
PL Genomics 25:538-546(1995).
RN [2]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE (LONG AND TRUNCATED FORMS).
RC Tissue=Brain;
EX MEDLINE=95123473; PubMed=7923156;
RA Shelton D.L.; Sutherland J.; Gripp J.; Camerato T.; Armanini M.P.;
RA Phillips H.S.; Carroll K.; Spencer S.D.; Levinson A.D.;
RT "Human Trks: molecular cloning, tissue distribution, and expression
RT of extracellular domain immunoadhesins.";
PL J. Neurosci. 15:477-491(1995)
RN [3]
RP SEQUENCE FROM N.A. (TRUNCATED FORMS).
RC Tissue=Hippocampus;
EX MEDLINE=9522162; PubMed 7336202;
FA Allen S.J.; Dabwani E.; Eckford S.D.; Wilce A.M.; Ashford M.;

```

RA Colebroock-Savory, Poonsey, P., Hargreaves, S.H.M.
RT Relocation of a putatively novel form of human TRP and detection of
R1 Neurexins and Neuroligins in human brain,
CC Neuroscience 60:825-834(1994).
CC - FUNCTION RECEPTOR FOR BRAIN-DERIVED NEUROTROPHIC FACTOR (BDNF),
CC NEUREGULIN 1 AND NEUREGULIN 4/5 BUT NOT NERVE GROWTH FACTOR
CC (NGF) INVOLVED IN THE DEVELOPMENT AND/OR MAINTENANCE OF THE
CC NERVOUS SYSTEM. THIS IS A TRANSMEMBRANE PROTEIN KINASE RELATED TO
CC CATERPILLAR FOR THE TRK RECEPTORS AND SHC, IT'S KINASE, AND ITS
CC CANVAS.
CC - CATALYTIC ACTIVITY: ATP + 3 protein tyrosine AMP + pyrophosphate
CC SUBSTRATE: PHOSPHATASE.
CC - AFFINITY AND SPECIFIC HIGH AFFINITY BINDING SITES (NO SIMILARITY).
CC - REGULATION: ACTIVATION BY BDNF AND NGF.
CC - ALTERNATIVE SPLICING: A TRANSCRIPT WITH AN ALTERNATIVE
CC EXON THAT REMAINS SPREAD THROUGHOUT THE TISSUE FROM IS EXPRESSED IN
CC FETAL TISSUE IN ADULT BRAIN.
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED, MAINLY IN THE NERVOUS
CC TISSUE. IN THE CNS, EXPRESSION IS OBSERVED IN THE CEREBRAL CORTEX,
CC HIPPOCAMPS, THALAMUS, TRIGEMINOTRIBUTARY NERVE, SPINAL LAYER OF THE
CC SPERMATOGENESIS, BRAIN STEM, AND SPINAL CORD. IN THE PERIPHERAL NERVOUS
CC SYSTEM, IT IS EXPRESSED IN MANY TYPICAL SENSORY, THE MOTOR
CC NERVE, THE VESTIBULAR COCHLEAR NERVE, MULTIPLE FACIAL STRUCTURES, THE
CC SEMI-CIRCULAR CANALS, AND TESTIS ECTOPY.
CC - PTM: DIANT MEDIATED AUTO-PHOSPHORYLATION.
CC - SIMILARITY: BELONGS TO THE TRK FAMILY OF PROTEIN KINASES. MULTIN
CC RECEPTOR SUBFAMILY.
CC - SIMILARITY: CONTAINS A SERINE RICH REPEATS (SER)
CC - SIMILARITY: CONTAINS A TRANSDUCEDIN LIKE OF THE DOMAINS.
CC - SWISS PROT entry is available. It is predicted the gl...
CC between the Swiss Institute of Bioinformatics and the EMBL database.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified or altered, provided it is properly cited by email to
CC authors to inform a license agreement (see http://www.ebi.ac.uk/submit.html/
CC or send an email to license@ebi.ac.uk).

DR EMBL: 572140; AAC51371.1;
DR EMBL: 574473; AAH3109.1;
DR EMBL: 576474; AAB3110.1;
DR EMBL: X75358; CAAC3571.1;
DR HSPB: P06213; IIRK.
DR GenBank: U00018032; NTRK2.
DR MIM: 600456;
DR InterPro: IP000719; Euk_kinase.
DR InterPro: IP0003006; Ig_MHC.
DR InterPro: IP003598; Ig_C2.
DR InterPro: IP001611; DRK.
DR InterPro: IP000463; LRR_Octm.
DR InterPro: IP000372; LRR_Nectm.
DR InterPro: IP0002011; RTN4-NT1.
DR InterPro: IP001245; Ty1_kinase.
DR Pfam: PF000347; Ig_1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00069; LRR_1.
DR Pfam: PF01462; LRRT; 1.
DR Pfam: PF01463; LRRC; 1.
DR PRINTS: PR00103; TYRKINSO.
DR PROSITE: PS000001; Euk_kinase; 1.
DR SMART: SM00408; IG02; 1.
DR SMART: SM00082; LRRC; 1.
DR SMART: SM00013; LRRT; 1.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00167; PROTEIN_KINASE_ATD; 1.
DR PROSITE: PS00168; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00169; KINETIC_GTP_BIND; 1.
DR PROSITE: PS00170; PROTEIN_KINASE_TYR; 1.
DR Transmembrane, Type IV, Protein Kinase, Transmembrane, ATD Binding,
DR Transmembrane, Type IV, Protein Kinase, Transmembrane, Signal,
DR Transmembrane, Type IV, Protein Kinase, Transmembrane, Signal,
DR Transmembrane, Type IV, Protein Kinase, Transmembrane, Signal,

```

PW  information (splitting); Polymorphism.
FT 01-CHAIN 1 31 ELENENT 1 WITH 4410RS RECEPTOR.
H1 01-CHAIN 32 948 HAIRACELLULAR (POTENTIAL).
H1 01-CHAIN 32 449 POTENTIAL.
H1 01-CHAIN 431 454 POTENTIAL.
H1 01-CHAIN 459 511 CYTOSOLIC (POTENTIAL).
H1 01-CHAIN 511 524 IPR 1.
H1 01-CHAIN 524 537 IPR 2.
H1 01-CHAIN 537 544 IPR 3.
H1 01-CHAIN 544 552 IPR 4.
H1 01-CHAIN 552 560 IPR 5.
H1 01-CHAIN 560 568 IPR 6.
H1 01-CHAIN 568 576 IPR 7.
H1 01-CHAIN 576 584 IPR 8.
H1 01-CHAIN 584 592 IPR 9.
H1 01-CHAIN 592 600 IPR 10.
H1 01-CHAIN 600 608 IPR 11.
H1 01-CHAIN 608 616 IPR 12.
H1 01-CHAIN 616 624 IPR 13.
H1 01-CHAIN 624 632 IPR 14.
H1 01-CHAIN 632 640 IPR 15.
H1 01-CHAIN 640 648 IPR 16.
H1 01-CHAIN 648 656 IPR 17.
H1 01-CHAIN 656 664 IPR 18.
H1 01-CHAIN 664 672 IPR 19.
H1 01-CHAIN 672 680 IPR 20.
H1 01-CHAIN 680 688 IPR 21.
H1 01-CHAIN 688 696 IPR 22.
H1 01-CHAIN 696 704 IPR 23.
H1 01-CHAIN 704 712 IPR 24.
H1 01-CHAIN 712 720 IPR 25.
H1 01-CHAIN 720 728 IPR 26.
H1 01-CHAIN 728 736 IPR 27.
H1 01-CHAIN 736 744 IPR 28.
H1 01-CHAIN 744 752 IPR 29.
H1 01-CHAIN 752 760 IPR 30.
H1 01-CHAIN 760 768 IPR 31.
H1 01-CHAIN 768 776 IPR 32.
H1 01-CHAIN 776 784 IPR 33.
H1 01-CHAIN 784 792 IPR 34.
H1 01-CHAIN 792 800 IPR 35.
H1 01-CHAIN 800 808 IPR 36.
H1 01-CHAIN 808 816 IPR 37.
H1 01-CHAIN 816 824 IPR 38.
H1 01-CHAIN 824 832 IPR 39.
H1 01-CHAIN 832 840 IPR 40.
H1 01-CHAIN 840 848 IPR 41.
H1 01-CHAIN 848 856 IPR 42.
H1 01-CHAIN 856 864 IPR 43.
H1 01-CHAIN 864 872 IPR 44.
H1 01-CHAIN 872 880 IPR 45.
H1 01-CHAIN 880 888 IPR 46.
H1 01-CHAIN 888 896 IPR 47.
H1 01-CHAIN 896 904 IPR 48.
H1 01-CHAIN 904 912 IPR 49.
H1 01-CHAIN 912 920 IPR 50.
H1 01-CHAIN 920 928 IPR 51.
H1 01-CHAIN 928 936 IPR 52.
H1 01-CHAIN 936 944 IPR 53.
H1 01-CHAIN 944 952 IPR 54.
H1 01-CHAIN 952 960 IPR 55.
H1 01-CHAIN 960 968 IPR 56.
H1 01-CHAIN 968 976 IPR 57.
H1 01-CHAIN 976 984 IPR 58.
H1 01-CHAIN 984 992 IPR 59.
H1 01-CHAIN 992 1000 IPR 60.
H1 01-CHAIN 1000 1008 IPR 61.
H1 01-CHAIN 1008 1016 IPR 62.
H1 01-CHAIN 1016 1024 IPR 63.
H1 01-CHAIN 1024 1032 IPR 64.
H1 01-CHAIN 1032 1040 IPR 65.
H1 01-CHAIN 1040 1048 IPR 66.
H1 01-CHAIN 1048 1056 IPR 67.
H1 01-CHAIN 1056 1064 IPR 68.
H1 01-CHAIN 1064 1072 IPR 69.
H1 01-CHAIN 1072 1080 IPR 70.
H1 01-CHAIN 1080 1088 IPR 71.
H1 01-CHAIN 1088 1096 IPR 72.
H1 01-CHAIN 1096 1104 IPR 73.
H1 01-CHAIN 1104 1112 IPR 74.
H1 01-CHAIN 1112 1120 IPR 75.
H1 01-CHAIN 1120 1128 IPR 76.
H1 01-CHAIN 1128 1136 IPR 77.
H1 01-CHAIN 1136 1144 IPR 78.
H1 01-CHAIN 1144 1152 IPR 79.
H1 01-CHAIN 1152 1160 IPR 80.
H1 01-CHAIN 1160 1168 IPR 81.
H1 01-CHAIN 1168 1176 IPR 82.
H1 01-CHAIN 1176 1184 IPR 83.
H1 01-CHAIN 1184 1192 IPR 84.
H1 01-CHAIN 1192 1200 IPR 85.
H1 01-CHAIN 1200 1208 IPR 86.
H1 01-CHAIN 1208 1216 IPR 87.
H1 01-CHAIN 1216 1224 IPR 88.
H1 01-CHAIN 1224 1232 IPR 89.
H1 01-CHAIN 1232 1240 IPR 90.
H1 01-CHAIN 1240 1248 IPR 91.
H1 01-CHAIN 1248 1256 IPR 92.
H1 01-CHAIN 1256 1264 IPR 93.
H1 01-CHAIN 1264 1272 IPR 94.
H1 01-CHAIN 1272 1280 IPR 95.
H1 01-CHAIN 1280 1288 IPR 96.
H1 01-CHAIN 1288 1296 IPR 97.
H1 01-CHAIN 1296 1304 IPR 98.
H1 01-CHAIN 1304 1312 IPR 99.
H1 01-CHAIN 1312 1320 IPR 100.
H1 01-CHAIN 1320 1328 IPR 101.
H1 01-CHAIN 1328 1336 IPR 102.
H1 01-CHAIN 1336 1344 IPR 103.
H1 01-CHAIN 1344 1352 IPR 104.
H1 01-CHAIN 1352 1360 IPR 105.
H1 01-CHAIN 1360 1368 IPR 106.
H1 01-CHAIN 1368 1376 IPR 107.
H1 01-CHAIN 1376 1384 IPR 108.
H1 01-CHAIN 1384 1392 IPR 109.
H1 01-CHAIN 1392 1400 IPR 110.
H1 01-CHAIN 1400 1408 IPR 111.
H1 01-CHAIN 1408 1416 IPR 112.
H1 01-CHAIN 1416 1424 IPR 113.
H1 01-CHAIN 1424 1432 IPR 114.
H1 01-CHAIN 1432 1440 IPR 115.
H1 01-CHAIN 1440 1448 IPR 116.
H1 01-CHAIN 1448 1456 IPR 117.
H1 01-CHAIN 1456 1464 IPR 118.
H1 01-CHAIN 1464 1472 IPR 119.
H1 01-CHAIN 1472 1480 IPR 120.
H1 01-CHAIN 1480 1488 IPR 121.
H1 01-CHAIN 1488 1496 IPR 122.
H1 01-CHAIN 1496 1504 IPR 123.
H1 01-CHAIN 1504 1512 IPR 124.
H1 01-CHAIN 1512 1520 IPR 125.
H1 01-CHAIN 1520 1528 IPR 126.
H1 01-CHAIN 1528 1536 IPR 127.
H1 01-CHAIN 1536 1544 IPR 128.
H1 01-CHAIN 1544 1552 IPR 129.
H1 01-CHAIN 1552 1560 IPR 130.
H1 01-CHAIN 1560 1568 IPR 131.
H1 01-CHAIN 1568 1576 IPR 132.
H1 01-CHAIN 1576 1584 IPR 133.
H1 01-CHAIN 1584 1592 IPR 134.
H1 01-CHAIN 1592 1600 IPR 135.
H1 01-CHAIN 1600 1608 IPR 136.
H1 01-CHAIN 1608 1616 IPR 137.
H1 01-CHAIN 1616 1624 IPR 138.
H1 01-CHAIN 1624 1632 IPR 139.
H1 01-CHAIN 1632 1640 IPR 140.
H1 01-CHAIN 1640 1648 IPR 141.
H1 01-CHAIN 1648 1656 IPR 142.
H1 01-CHAIN 1656 1664 IPR 143.
H1 01-CHAIN 1664 1672 IPR 144.
H1 01-CHAIN 1672 1680 IPR 145.
H1 01-CHAIN 1680 1688 IPR 146.
H1 01-CHAIN 1688 1696 IPR 147.
H1 01-CHAIN 1696 1704 IPR 148.
H1 01-CHAIN 1704 1712 IPR 149.
H1 01-CHAIN 1712 1720 IPR 150.
H1 01-CHAIN 1720 1728 IPR 151.
H1 01-CHAIN 1728 1736 IPR 152.
H1 01-CHAIN 1736 1744 IPR 153.
H1 01-CHAIN 1744 1752 IPR 154.
H1 01-CHAIN 1752 1760 IPR 155.
H1 01-CHAIN 1760 1768 IPR 156.
H1 01-CHAIN 1768 1776 IPR 157.
H1 01-CHAIN 1776 1784 IPR 158.
H1 01-CHAIN 1784 1792 IPR 159.
H1 01-CHAIN 1792 1800 IPR 160.
H1 01-CHAIN 1800 1808 IPR 161.
H1 01-CHAIN 1808 1816 IPR 162.
H1 01-CHAIN 1816 1824 IPR 163.
H1 01-CHAIN 1824 1832 IPR 164.
H1 01-CHAIN 1832 1840 IPR 165.
H1 01-CHAIN 1840 1848 IPR 166.
H1 01-CHAIN 1848 1856 IPR 167.
H1 01-CHAIN 1856 186
```

CC THE SECOND FOURTEENTH IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC ACTS, IMPROVING CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC TO USE PERPOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC -1- PATHWAY: Fattyunsaturated fatty acid biosynthesis
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLY)
CC -1- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/OR BE INVOLVED IN METAL ION BINDING
CC -1- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ibt.ch/announcements/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L29215; AAA50158.1;
DR InterPro: IPR001225; FA_desaturase.
DR Pfam: PF00487; FA_desaturase; 1.
DR ProDom: ProDom1991; FA_desaturase; 1.
KW Oxy-reductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW Transit peptide.
FT TRANSIT 1 63 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 64 424 OMEGA-6 FATTY ACID DESATURASE.
FT DOMAIN 165 169 HISTIDINE BOX-1.
FT DOMAIN 201 205 HISTIDINE BOX-2.
FT DOMAIN 361 365 HISTIDINE BOX-3.
SQ SEQUENCE 424 AA; 49641 MW; 4F3DF52B4B1A2009 CRC64;

Query Match 80.0%; Score 36; DB 1; Length 424;
Best Local Similarity 66.7%; Freq. No. 30;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVRWHF 6
Db 251 WLMWHF 256

RESULT 7
BCSB_XANAC STANDARD; PRT; 788 AA.
ID_BCSB_XANAC
AC P58933;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN 2002 (Rel. 41, Last sequence update)
DT 15-JUN 2002 (Rel. 41, Last annotation update)
DE Cyclic di-GMP binding protein precursor (cellulose synthase regulatory
DE subunit).
GN BCSB OR XAC3517;
OC Xanthomonas axonopodis (p. strain).
OC Bacteria: Proteobacteria: gamma subphylum: Xanthomonas group.
OC Xanthomonas.
OX NCBI_TaxID:92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=2024217;
RA da Silva A C P, Perrot A, Polinac F C, Farth C S, Forlan L F,
RA Onagdio R R, Monteiro-Vitorello C B, Van Sluys W A, Almeida N F,
RA Alves I M C, de Amiral A M, Bertolini M C, Camargo L B A,
RA Camarotte G, Cannavan F, Cardoso J, Chatterjee F, Clafina L F,
RA Ciarrelli P M P, Coutinho L I, Cristina Santos J F, Florschütz H,
RA Faria J B, Ferreira A T S, Ferreira P C C, Ferro M I T,
RA Formighieri F F, Franco M C, Greggio C C, Gruber A,
RA Katsuyama A M, Kishi I T, Leite P P, Lemos F G M, Lemos M V F,
RA Locati E C, Machado W A, Madeira A M B R, Martinez-Rosa N M,
RA Martins E C, Melandri J, Monck C F M, Miyaki G Y, Moran P H,
RA Moreira L M, New W T M, Okada V F, Oliveira M C, Oliveira V P,
RA Pereira H A, Rossi A, Seta A A D, Silva G, de Souza R F,
RA Spinola R A F, Takita M A, Tamura P F, Teixeira F C, Torres P I D,
RA Trindade dos Santos M, Truffi D, Tsai S M, White F F,
RA Setubal J C, Kitajima J P;

FT "Comparison of the genomes of two Xanthomonas pathogens with differing
FT host specificities".
FT Nature 417:459-463(2002).
CC -1- FUNCTION: Binds the cellulose synthase activator, bcsA (3-5').
CC -1- Cyclic di-guanylic acid (c-di-GMP) (by similarity).
CC -1- SEQUENCE: tightly associated with the cellulose synthase catalytic
CC subunit (by similarity).
CC -1- PATHWAY: bacterial cellulose biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE ACSB/BCSB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ibt.ch/announcements/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE012003; AAM38360.1; ALT_INIT.
KW Cellulose biosynthesis; Signal.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 788 CYCLIC DI-GMP BINDING PROTEIN.
SQ SEQUENCE 788 AA; 85117 MW; 067ACCF11F155070F04;
Query Match 80.0%; Score 36; DB 1; Length 788;
Best Local Similarity 66.7%; Freq. No. 56;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WVRWHF 6
Db 743 WLMWHF 748

RESULT 8
SUV3_DROME STANDARD; PRT; 1169 AA.
ID_SUV3_DROME
AC P20193;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Suppressor of variegation protein 3-7.
GN SUVAR(3)7.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9523788; PubMed=7708405;
RA Cleard F, Matsarskaia M, Spierer P;
FT "The modifier of position effect variegation (SUV3) of Drosophila:
FT there are two alternative transcripts and seven scattered zinc
FT fingers, each preceded by a tryptophan box".
RL Nucleic Acids Res 23:796-802(1995).
RN [2]
RP ERRATUM.
RA Cleard F, Matsarskaia M, Spierer P;
FT Nucleic Acids Res 23:3904-3804(1995).
RN [3]
RX MEDLINE=951169 FROM N.A.
RX MEDLINE=20190836; PubMed 2107402;
FA Feeder G, Glazier M, Farth C, Jauz G, Spierer A, Spierer P;
FT "Dependence of position effect variegation in Drosophila on level of a
FT gene encoding an unusual zinc finger protein".
FL Nature 344:219-223(1990).
CC -1- FUNCTION: THIS PROTEIN IS A POSITIVE FACTOR IN POSITION-
CC EFFECT VARIATION, THE INACTIVATION IN SOME CELLS OF A GENE
CC TRANSLOCATED NEXT TO HETEROCHROMATIN. IT COULD PLAY A ROLE IN
CC CHROMOSOME CONDENSATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

DB 178 WCNWHF 183

RESULT 11

RS12_METJA STANDARD, FRT, 148 AA.
 ID RS12_METJA
 AC P54062;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 30S ribosomal protein S12P.
 GN RPS12P OR M01046
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales;
 OC Methanocaldococcus; Methanocaldococcus
 OX NCBI_TaxID=2100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN: ATCC 35061; ATCC 35061;
 RX MEDLINE=96337009, PubMed 8688987,
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., White O., Fitzgerald L.M., Clayton P.A., Gocayne J.D.,
 RA Kerlavage A.P., Dougherty P.A., Tomb L.F., Adams M.D., Betcher C.T.,
 RA Ouellette P., Whitehead E.F., Weinstock P.G., Motro J.M., Olsen A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fierman J.T., Nguyen D.,
 RA Uitterlinden P., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.W., Harst M.A., Kalne R.P., Parodovsky M.,
 RA Klenk H.P., Fraser C.M., Smith H.C., Weese C.P., Venter J.C.,
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii";
 RL Science 273:1058-1073 (1996)
 CC FUNCTION: PROTEIN S12P IS INVOLVED IN THE TRANSLATION INITIATION
 CC STEP (BY SIMILARITY).
 CC SIMILARITY BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/annotation/>
 CC or send an email to license@isb.sib.ch)
 CC
 CC EMBL, U07547, AAB00500.1,
 CC TIGR, M01046;
 CC InterPro: IPR000230; Ribosomal_S12
 CC Pfam: PF00164; Ribosomal_S12; 1
 CC ProDom: PD000576; Ribosomal_S12; 1
 CC TIGRFAMs: TIGR00982; S23_S12_E_A; 1.
 CC PROSITE: PS00065; RIBOSOMAL_S12; 1.
 CC KW Ribosomal protein; Complete proteome
 CC SEQUENCE 148 AA; 16494 MW; 424174480Da; 9990 SpA4;
 CC
 CC Query Match 75.6%; Score 34; DE 1; Length 148;
 CC Best Local Similarity 90.0%; Pred. No. 23;
 CC Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 WVRWH 5
 CC DB 22 WCRWH 26

RESULT 12

CBL_ECOLI STANDARD, FRT, 316 AA.
 ID CBL_ECOLI
 AC Q47083; P76353;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcriptional regulator cbl.
 GN CBL OR B1987.
 OS Escherichia coli.

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC OX NCBI_TaxID=562;
 CC RN [1]
 CC SEQUENCE FROM N.A.
 CC RP STRAIN-K12;
 CC RX MEDLINE=96105196; PubMed 8529872;
 CC PA Iwanicka-Nowicka R., Hryniewicz M.M.;
 CC RT "A new gene, cbl, encoding a member of the LysR family of
 CC PT transcriptional regulators belongs to Escherichia coli cys regulation";
 CC PL Gene 166:11-17 (1995).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-K12;
 CC RX MEDLINE=97426617; PubMed 9278503;
 CC PA Plattner F.P., Plunkett G. III, Pluck G.A., Patra N.T., Pauland V.,
 CC PA Riley M., Collins-Vides I., Gasser J.D., Podge C.P., Mayhew G.F.,
 CC RA Greger T., Davis N.W., Kilypatov H.A., Gaudin M.A., Rose D.C.,
 CC RA Mau B., Shao Y.;
 CC RC "The complete genome sequence of Escherichia coli K-12";
 CC RL Science 277:1233-1238 (1997).
 CC RN [3]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-K12;
 CC RX MEDLINE=97051358; PubMed 9090040;
 CC PA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 CC PA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 CC PA Makino F., Miki T., Mizobuchi K., Mori H., Moti T., Moriyama K.,
 CC PA Nakade S., Nakamura Y., Nishimoto H., Nishie Y., Oshima T.,
 CC PA Saito N., Sawai G., Saki Y., Sivasubramanian S., Tagami H.,
 CC PA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 CC PT "A 4.6 Mb DNA sequence of the Escherichia coli K-12 genome
 CC FT corresponding to the 4.3-5.0 Mb region on the linkage map";
 CC RL DNA Res. 3:379-392 (1996).
 CC RN [4]
 CC RP FUNCTION: MAY BE AN ACTIVATOR REGULATOR PROTEIN WITHIN THE CYS
 CC REGULATION.
 CC
 CC SIMILARITY BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/annotation/>
 CC or send an email to license@isb.sib.ch)
 CC
 CC EMBL, U31629; AAA99229.1;
 CC FMRP; AB000290; AAC75049.1;
 CC EMBL; D90837; BAA15805.1;
 CC DR HSSP; P45600; IAL3.
 CC DR EcGene; E014264; CBL.
 CC DR InterPro: IPR000847; HTH_LysR.
 CC DR InterPro: IPR005119; LysR_subst
 CC PFam; PF00126; HTH_1; 1.
 CC PFam; PF03466; LysR_substrate; 1.
 CC PROSITE: PS00039; HTH_LysR.
 CC DR PROSITE; PS00044; HTH_LysR_FAMILY; 1.
 CC KW Transcription regulation; DNA-binding; Complete proteome.
 CC FT DNA BIND 19 39 H-T-H MOTIF (POTENTIAL).
 CC FT CONFLICT 143 143 A > R (IN REF. 1).
 CC SEQUENCE 316 AA; 35856 MW; 700R594765010670 CRC64;

RESULT 13

QY 1 WVRWH 5
 DB 163 WFRWH 167

RESULT 13


```
CBL_KLEAF
ID CBL_KLEAF STANDARD; PRT; 316 AA.
AC CCR598;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Transcriptional regulator cbl.
GN CBL.
OS Homo sapiens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Klebsiella.
OX NCBI TaxID=28481.
RN
RP SEQUENCE FROM N.A.
RC STRAIN W72 / K01243.
RX MEDLINE 3323337, PubMed 5458833.
RA Schwacha A., Bender P.A.
RT "The cblA gene encodes a protein that is a putative histone-like
  aerogenes."
RL 01. Bacteriol. 175:2107-2115(1993).
RC
CC -1- FUNCTION. MAY BE AN ACCESSORY REGULATOR PROTEIN WITHIN THE CBL
  REGION.
CC -1- SIMILARITY. BELONGS TO THE LYSR FAMILY OF TRANSGLUTAMINASE
  REGULATORS.
CC
CC This SWISS PROTEIN entry is linked through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL database to
  the European Bioinformatics Institute. These two databases contain the
  use by scientists. Information is being brought to you in a way
  modified and this is done in a way that is not intended to be
  entities requires a license agreement (see http://www.ebi.ac.uk/EMBL/
  or send an email to license@ebi.ac.uk).
```

```
PA Zweiger G., Marczyński G., Shapiro L.
FI "A Caulobacter DNA methyltransferase that functions only in the
  predivisioal cell."
RL J. Mol. Biol. 235:472-485(1994).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CBL5.
RX MEDLINE 2173099, PubMed 11259647.
RA Mouton R., Polgar J., Mack M.T., Paulson L.T., Nelson K.E.,
  Eisen D., Helbergh J.E., Alley M.E.R., Ohta N., Madock J.R.,
  Folsch U., Nelson W.C., Haxson A., Stephens C., Clarke M.D., Fry B.,
  DeBoy R.T., Dodson R.D., Durkin A.S., Gwinn M.L., Haft D.H.,
  Kutylov I.F., Olt G., Orren M.P., Paturi H., Shetty J., Berry K.,
  Delubek T., Tian K., Rife A., Wamboldt J., Erogluova M., White O.,
  Salzberg S.L., Venter D.C., Shapiro L., Fraser C.M.
RI "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC
CC HIGHLIGHT: "THE METHYLTRANSFERASE THAT ENCODES THE CBLA GENE IS
  ANO AND CDS233 SPECIFIC METHYLATION ON A-D ON BOTH STRANDS.
  CDS233 AND CDS234 SPECIFIC METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
  ASSESS TO CONTRIBUTE TO THE ACCURATE CELL-CYCLE CONTROL OF DNA
  REPLICATION AND CELLULAR MORPHOLOGY.
  1. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + DNA adenine 3'-
    adenosyl-L-histidine + RNA 6-methyladenosine.
  2. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  3. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  4. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  5. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  6. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  7. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  8. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  9. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  10. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  11. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  12. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  13. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  14. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  15. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  16. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  17. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  18. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  19. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  20. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  21. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  22. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  23. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  24. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  25. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  26. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  27. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  28. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  29. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  30. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  31. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  32. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  33. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  34. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  35. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  36. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  37. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  38. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  39. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  40. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  41. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  42. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  43. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  44. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  45. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  46. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  47. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  48. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  49. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  50. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  51. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  52. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  53. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  54. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  55. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  56. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  57. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  58. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  59. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  60. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  61. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  62. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  63. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  64. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  65. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  66. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  67. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  68. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  69. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  70. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  71. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  72. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  73. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  74. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  75. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  76. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  77. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  78. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  79. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  80. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  81. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  82. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  83. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  84. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  85. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  86. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  87. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  88. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  89. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  90. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  91. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  92. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  93. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  94. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  95. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  96. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  97. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  98. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  99. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
  100. ACTIVITY ACTIVITY: 3-adenosyl-L-histidine + RNA 6-methyladenosine.
```

```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:99156852; PubMed 10036181;
RA Marchese A., Sawzdargo M., Nguyen T., Cheng R., Heng H.H., Nowak T.,
RA Im D.S., Lynch K.P., George S.R., O'Dowd B.F.;
RT "Discovery of three novel orphan G-protein-coupled receptors";
RL Genomics 56:12-21(1999).
CC -!- FUNCTION: Orphan receptor. May play a role in brain function.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in brain; detected in the basal
CC forebrain, frontal cortex, and caudate, but not in thalamus,
CC hippocampus, or putamen.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This swiss-prot entry is copyright It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF118266; AAD21056.1;
DR Genew; HGNC:4503; GPR45.
DR MIM: 604838;
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHDDPSN.
DR PROSITE; PS00517; G_PROTEIN_RECEP_F1_1, FALSE_NEG.
DR PROSITE; PSS07622; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor, Transmembrane, Glycoprotein.
FT DOMAIN 1 38 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 39 59 1 (POTENTIAL).
FT DOMAIN 60 75 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 76 96 2 (POTENTIAL).
FT DOMAIN 97 109 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 110 130 3 (POTENTIAL).
FT DOMAIN 131 149 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 150 170 4 (POTENTIAL).
FT DOMAIN 171 198 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 199 219 5 (POTENTIAL).
FT DOMAIN 220 268 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 269 289 6 (POTENTIAL).
FT DOMAIN 290 305 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 306 326 7 (POTENTIAL).
FT DOMAIN 327 372 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 4 4 N-LINKED (GLCNAC:...) (POTENTIAL).
FT CARBOHYD 17 17 N-LINKED (GLCNAC:...) (POTENTIAL).
FT CARBOHYD 20 20 N-LINKED (GLCNAC:...) (POTENTIAL).
SQ SEQUENCE 372 AA; 42000 MW; R4R4FQ7AEDE20199 CRC64;
Query Match 75.6%; Score 34; DB 1; Length 372;
Best Local Similarity 100.0%; Pred.No. 57;
Matches 0; Conservativa 0; Mismatches 0; Indels 0; Gaps 0;
```

Search completed: February 20, 2003, 10:48:01
Job time : 8.66667 secs



GenCore version 5.1.1
Copyright 1999-2002 by GenCore, Inc.

OM protein Protein search, using SW-FAST

Run on: February 21, 2003, 10:45:10, Search time 23.333 seconds
(without alignments)
52.984 Million cell updates/sec

Title: US-09-912-414-2

Perfect score: 45

Sequence: 1 WVRWHP 6

Scoring table: BLOSUM62GX

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 286647116 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Part: Pseudomonas, Minimum Match: 0%

Maximum Match: 100%

Listing first 45 summaries

Database:

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	91.1	219	10 Q94K47	Q94K47 arabidopsis
2	41	91.1	256	10 Q94K12	Q94K12 arabidopsis
3	40	88.9	201	10 Q94K50	Q94K50 arabidopsis
4	40	88.9	202	10 Q94XY3	Q94XY3 brassica ra
5	40	88.9	213	10 Q94XY5	Q94XY5 brassica ra
6	40	88.9	216	10 Q94XY2	Q94XY2 brassica ra
7	40	88.9	222	10 Q94XY4	Q94XY4 brassica ol
8	39	86.7	148	16 Q94D55	Q94D55 streptomyces
9	39	86.7	293	16 Q94D65	Q94D65 pseudomonas
10	39	86.7	308	6 Q94WV6	Q94WV6 capra hircu
11	39	86.7	370	10 Q94W32	Q94W32 arabidopsis
12	39	86.7	407	10 Q94W34	Q94W34 arabidopsis
13	38	84.4	70	4 Q8WXJ4	Q8WXJ4 homo sapien
14	38	84.4	178	6 Q94K65	Q94K65 macaca fasc
15	38	84.4	537	4 Q8WXJ6	Q8WXJ6 homo sapien
16	38	84.4	553	4 Q8WXJ5	Q8WXJ5 homo sapien

17	38	84.4	838	4 Q8WXJ7	Q8WXJ7 homo sapien
18	37	82.2	591	4 Q13041	Q13041 homo sapien
19	36	80.0	185	16 Q94K51	Q94K51 pseudomonas
20	36	80.0	223	16 Q94K24	Q94K24 pseudomonas
21	36	80.0	430	16 Q94K54	Q94K54 pseudomonas
22	36	80.0	462	16 Q94K53	Q94K53 pseudomonas
23	36	80.0	464	16 Q94K25	Q94K25 pseudomonas
24	36	80.0	465	16 Q94K26	Q94K26 pseudomonas
25	36	80.0	466	16 Q94K27	Q94K27 pseudomonas
26	36	80.0	467	16 Q94K28	Q94K28 pseudomonas
27	36	80.0	468	16 Q94K29	Q94K29 pseudomonas
28	36	80.0	469	16 Q94K30	Q94K30 pseudomonas
29	36	80.0	470	16 Q94K31	Q94K31 pseudomonas
30	36	80.0	471	16 Q94K32	Q94K32 pseudomonas
31	36	80.0	472	16 Q94K33	Q94K33 pseudomonas
32	36	80.0	473	16 Q94K34	Q94K34 pseudomonas
33	36	80.0	474	16 Q94K35	Q94K35 pseudomonas
34	36	80.0	475	16 Q94K36	Q94K36 pseudomonas
35	36	80.0	476	16 Q94K37	Q94K37 pseudomonas
36	36	80.0	477	16 Q94K38	Q94K38 pseudomonas
37	36	80.0	478	16 Q94K39	Q94K39 pseudomonas
38	36	80.0	479	16 Q94K40	Q94K40 pseudomonas
39	36	80.0	480	16 Q94K41	Q94K41 pseudomonas
40	36	80.0	481	16 Q94K42	Q94K42 pseudomonas
41	36	80.0	482	16 Q94K43	Q94K43 pseudomonas
42	36	80.0	483	16 Q94K44	Q94K44 pseudomonas
43	36	80.0	484	16 Q94K45	Q94K45 pseudomonas
44	36	80.0	485	16 Q94K46	Q94K46 pseudomonas
45	36	80.0	486	16 Q94K47	Q94K47 pseudomonas

ALIGNMENTS

RESULT 1

Q94K47 PPTREMBL_21: 219 AA.
AC Q94K47: PPTREMBL_21: 219 AA.
DT 01-DEC-2001 (TREMBL_19, Created)
DT 01-DEC-2001 (TREMBL_19, Last sequence update)
RT 01-JUN-2002 (TREMBL_21, Last annotation update)
DE Putative 6-phosphogluconolactonase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
PA Yamada K., Liu S.X., Sakano H., Pham P.K., Ban H., Chung M.K.,
FA Gollub A.B., Lee T.M., Smith H.T., Tang Y., Tabor C.,
PA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida O.,
PA Jones T., Kamiya A., Karlin-Neumann G., Kawai T., Kim C., Kosemura E.,
PA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
PA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
PA Shinozaki K., Davis P.W., Ecker J.P., Theologis A.,
RT "Full Length cDNA of gene K16H17.11/At5g24400 (GI:10177901)."
PI Submitted (APR-2001) to the FMP/Genbank/DBP1 databases.
CC - SIMILARITY: CONTAINS 1 WD REPEAT (TRF ASP DOMAIN).
DP EMBL: AP030355; AAK44120.1; -
DR TIGR: TIGR01198; pg1; 1.
RW Repeat: WD repeat.
SQ SEQUENCE: 219 AA; 24542 MW; 12044723DC34348A GC664;

Query Match: 91.1%; Score 41; DB 10; Length 219;

Best Local Similarity: 83.3%; Pred. No. 23;

Matches: 5; Conservative: 0; Mismatches: 1; Indels: 0; Gaps: 0;

CY 1 WVRWHP 6
DB 72 WARWHP 77

```

RESULT 2
Q9FIN2      PRELIMINARY;      PRT;      256 AA.
ID Q9FIN2;
AC Q9FIN2;
DT 01-MAR-2001 (TREMblrel, 15, Created)
DT 01-MAR-2001 (TREMblrel, 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE 6 phosphogluconate dehydrogenase-like protein.
OS Arabidopsis thaliana (Mossberg et al)
OC Eukaryota; Viridiplantae; Embryophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosales;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=COLIMPIA;
RX MEDLINE=99156233; PubMed=10048486;
RA Araiwa T., Oate C., Kaneko T., Nakamura Y., Kozaki H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5, VIII.
RT Sequence features of the regions of 1,091,059 bp covered by seventeen
RT physically mapped B1 and TAC clones."
RL DNA Res. 3:373-391(1998).
DR EMBL; AB016061, pap10001.1,
DR InterPro; IPR000457, Glucosaminidase;
DR Pfam; PF01162, Glucosamine_1su; I.
DR TrEMBL; TIGR01198; P91; I.
SQ SEQUENCE 256 AA, 28034 MW, 2187046ECTDC072AF CRC64;

Query Match      89.9%, Score 40, DB 10, Length 201,
Best Local Similarity 83.3%, Pred. NO. 32,
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVMWHP 6
DB 72 WARMHP 77

RESULT 3
Q9C603      PRELIMINARY;      PRT;      201 AA.
ID Q9C603;
AC Q9C603;
DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE Omega 6 reductase (fragment).
OS Arabidopsis thaliana (Mossberg et al)
OC Eukaryota; Viridiplantae; Embryophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosales;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=CV; COLIMBIA;
RA Fourmann M., Proger N., Brunel D.;
RT "Amplified consensus gene markers: tools designing for a genetic map
RT of Arabidopsis known-function genes in Brassica."
RT Submitted (JAN 2000) to the EMBL/GenBank/TrEMBL databases.
RL EMBL; AF223368, AK000601.1,
DR InterPro; IPR001225; FA_desaturase;
DR Pfam; PF00487, FA_desaturase; I.
DR ProDom; PD001081, FA_desaturase; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 201 AA, 23064 MW, 4011905577100 CRC64;

Query Match      89.9%, Score 10, DB 10, Length 201,
Best Local Similarity 83.3%, Pred. NO. 32,
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVMWHP 6
DB 72 WARMHP 77

```

```

Query Match      89.9%, Score 10, DB 10, Length 201,
Best Local Similarity 83.3%, Pred. NO. 32,
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVMWHP 6
DB 72 WARMHP 77

```

```

DB 156 WVMWHP 161

RESULT 4
Q9AXY3      PRELIMINARY;      PRT;      202 AA.
ID Q9AXY3;
AC Q9AXY3;
DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE Omega 6 reductase (fragment).
OS Brassica campestris (Field mustard).
OC Eukaryota; Viridiplantae; Embryophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosales;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3711;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=CV; P100;
RA Fourmann M., Proger N., Brunel D.;
RT "Amplified consensus gene markers: tools designing for a genetic map
RT of Arabidopsis known-function genes in Brassica."
RL Submitted (JAN 2000) to the EMBL/GenBank/TrEMBL databases.
DR EMBL; AF223368, AK000601.1,
DR InterPro; IPR001225; FA_desaturase;
DR Pfam; PF00487, FA_desaturase; I.
DR ProDom; PD001081, FA_desaturase; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 202 AA, 23063 MW, 400433AAB1E000 CRC64;

Query Match      89.9%, Score 40, DB 10, Length 202,
Best Local Similarity 83.3%, Pred. NO. 30,
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WVMWHP 6
DB 157 WVMWHP 162

RESULT 5
Q9AXY5      PRELIMINARY;      PRT;      213 AA.
ID Q9AXY5;
AC Q9AXY5;
DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE Omega 6 reductase (fragment).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Embryophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosales;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=CV; STELLAR;
RA Fourmann M., Proger N., Brunel D.;
RT "Amplified consensus gene markers: tools designing for a genetic map
RT of Arabidopsis known-function genes in Brassica."
RL Submitted (JAN 2000) to the EMBL/GenBank/TrEMBL databases.
DR EMBL; AF223368, AK000601.1,
DR InterPro; IPR001225; FA_desaturase;
DR Pfam; PF00487, FA_desaturase; I.
DR ProDom; PD001081, FA_desaturase; 2.
FT NON_TER
FT NON_TER
SQ SEQUENCE 213 AA, 2471 MW, 11584935AK100 CRC64;

Query Match      89.9%, Score 40, DB 10, Length 213,
Best Local Similarity 83.3%, Pred. NO. 32,
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Query Match      89.9%, Score 40, DB 10, Length 213,
Best Local Similarity 83.3%, Pred. NO. 32,
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 WVRWHF 6
 |||||
 DB 154 WVRWHF 159

RESULT 6

Q9AXY2

ID Q9AXY2 PRELIMINARY; PRT; 216 AA.

AC Q9AXY2;

DT 01-JUN-2001 (TEMBLrel. 17, Created)

DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Omega 6 reductase (Fragment).

GN FAD6 BN 1.

OS Brassica napus (Rapeseed).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;

OC eucotyledons; Brassicales; Brassicaceae; Brassica.

OX NCBI_TaxID=1709;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=CV, STELLAR;

RA Fournier M., Froger N., Brunel D.;

RT "Amplified consensus gene markers: Tools designing for a genetic map

RT of Arabidopsis-known-function genes in Brassica";

RL Submitted (JAN 2000) to the EMBL/GenBank/CCPDB databases

DR EMBL; AF224391; AAF00643.1;

DR InterPro; IPR001225; FA_desaturase.

DR Pfam; PF00487; FA_desaturase; 1.

DR ProDom; PD01081; FA_desaturase; ?

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

FT NON_TER 1

Query Match 88.9%; Score 40; DB 10; Length 222;
 Best Local Similarity 83.3%; Pred. No. 33;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WVRWHF 6
 |||||
 DB 163 WVRWHF 168

RESULT 8

Q9RDS5

ID Q9RDS5 PRELIMINARY; PRT; 148 AA.

AC Q9RDS5;

DT 01-MAY-2000 (TEMBLrel. 13, Created)

DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Putative membrane protein.

GN SC02242 CP SC022.04C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomyces; Streptomyces; Streptomyces;

OX NCBI_TaxID=1902;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=43(2) / M145;

RA Bentley S.D., Chater K.F., Gerdeiro-Tarraga A.M., Challis G.L.,

RA Thomson N.P., James K.D., Harris P.E., Quail M.A., Kieser H.,

RA Harper D., Bawden A., Brown S., Chandra S., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Higgins T., Hornsby T., Howarth S.,

RA Huang C.-H., Kieser T., Lake J.A., Murphy L., Oliver K., O'Neill S.,

RA Parkin W.L.S., Saunders D., Sharp S., Squares R., Taylor K.,

RA Warren T., Witzorrek A., Woodward J., Barrell B.G., Parkhill J.,

FA "Complete genome sequence of the model actinomycete Streptomyces

FT coelicolor A3(2)";

RL Nature 417-141-147(2002).

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

DR EMBL; AL136500; CAP64376.1;

Query Match 96.7%, Score 20, DP 16, Length 203,
 Best Local Similarity 92.3%, Pred. No. 67,
 Matches 5, Conservative 0, Mismatches 1, Indels 0, Gaps 0

1 WVRWH 5
 101 WVRWH 106

RESULT 10
 ID Q9M932 PRELIMINARY, PRT, 308 AA
 AC Q9M932
 DT 01-MAR-2002 (TREMBlrel 20, Created)
 DT 01-MAR-2002 (TREMBlrel 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
 DE Putative phenolase [EC:1.11.1.37]F1.
 GN GPR1
 OS Capra hircus (Goat).
 OC Eukaryota, Metazoa, Chordata, Carnivora, Vertebrata, Euteleostomi,
 OC Mammalia, Eutheria, Cetartiodactyla, Ruminantia, Bovina, Bovidae,
 OC Bovidae, Caprinae, Capra
 OX NCBI_TaxID=9926;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SHIBA;
 FA Watanabe Y., Nishiyama Y., Takaki H., Nishii Y., Nishii Y.,
 FT "The first full-length cDNA for the EMPL/Donkank/PRP 4th phases
 RT of the 1st phase of the EMPL/Donkank/PRP 4th phases
 RL Submitted (2011001) to the EMBL/Genbank/PRP 4th phases
 DR EMBL:AB064661; EMBL:AB064661, 1
 DE InterPro: IPR001440; GPR1_Protein
 DR InterPro: IPR001440; GPR1_Protein
 DR Pfam: PF00400; WD40_7
 DR PROSITE: PS00320; GPR1_Protein
 KM Repeat: WD repeat
 ST SEQUENCE FROM N.A.

Query Match 96.7%, Score 20, DP 16, Length 203,
 Best Local Similarity 92.3%, Pred. No. 67,
 Matches 5, Conservative 0, Mismatches 1, Indels 0, Gaps 0

QY 1 WVRWH 5
 101 WVRWH 106

RESULT 11
 ID Q9M932 PRELIMINARY, PRT, 370 AA
 AC Q9M932
 DT 01-OCT-2000 (TREMBlrel 15, Created)
 DT 01-OCT-2000 (TREMBlrel 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
 DE F14023_1 Protein
 GN F14023_1
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 OC Spermatophyta, Magnoliophyta, Eudicotyledons, Core Eudicotyledons, Rosales,
 OC Euphorbiales, Brassicales, Brassicaceae, Arabidopsids
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;

FA Liu Q., Li S., Lee J., Sakano H., Hasegawa A., Deng C., Toriumi M.,
 FA Chin S., Shin S., Choi E., Gotoh T., Hwang E., Choi T., Li J.,
 FA Liu A., Shao F., Vaynsberg M., Alifan H., Fuchida T., Chao Q., Com L.,
 FA Conway A.B., Hansen N., Johnson-Hansen C., Khan S., Kim C., Lam B.,
 FA Nguyen M., Park C., Shin P., Tamura G., Davis P., Ecker J.,
 FA Rodriguez N., Theologis A.,
 FA "The sequence of BAC F14023 from Arabidopsis thaliana chromosome 1,"
 FT Submitted (00V100) to the EMBL/Genbank/PRP 4th phases.
 RL Submitted (00V100) to the EMBL/Genbank/PRP 4th phases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 FA Theologis A.,
 RL Submitted (MDP-2000) to the EMBL/Genbank/PRP 4th phases.
 DT 01-OCT-2000 (TREMBlrel 15, Created)
 DT 01-OCT-2000 (TREMBlrel 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
 DE F14023_1 Protein
 GN F14023_1
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 OC Spermatophyta, Magnoliophyta, Eudicotyledons, Core Eudicotyledons, Rosales,
 OC Euphorbiales, Brassicales, Brassicaceae, Arabidopsids
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;

Query Match 96.7%, Score 20, DP 16, Length 203,
 Best Local Similarity 92.3%, Pred. No. 67,
 Matches 5, Conservative 0, Mismatches 1, Indels 0, Gaps 0

QY 1 WVRWH 5
 151 WVRWH 155

RESULT 12
 ID Q9M932 PRELIMINARY, PRT, 407 AA
 AC Q9M932
 DT 01-JUN-2002 (TREMBlrel 21, Created)
 DT 01-JUN-2002 (TREMBlrel 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
 DE Hypothetical protein
 GN ATIG71840
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 OC Spermatophyta, Magnoliophyta, Eudicotyledons, Core Eudicotyledons, Rosales,
 OC Euphorbiales, Brassicales, Brassicaceae, Arabidopsids
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;

Query Match 96.7%, Score 20, DP 16, Length 203,
 Best Local Similarity 92.3%, Pred. No. 67,
 Matches 5, Conservative 0, Mismatches 1, Indels 0, Gaps 0

QY 1 WVRWH 5
 160 WVRWH 164

RESULT 13

Q8WXJ4
ID Q8WXJ4 PRELIMINARY; PRT; 70 AA.
AC Q8WXJ4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Neurotrophin receptor tyrosine kinase type 2 (Fragment).
GN NTRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-2165693; PubMed-11798182;
RA Stojilov P., Castren E., Stamm S.;
RT "Analysis of the Human TrkR Gene Genomic Organization Reveals Novel
TrkR Isoforms, Unusual Gene Length, and Splicing Mechanism."
RL Biochem Biophys Res Commun 240:104-106(2002).
DR EMBL; AF410902; AAL67968.1;
DR InterPro; IPR000372; IPR_Nterm
DR Pfam; PF01462; LRRNT; 1.
DR SMART; SM00013; LRRNT; 1.
KW Receptor; Kinase.
FT NON_TER 70
SQ SEQUENCE 70 AA; 744 MW; 1648443000B/23A CDD64;

Query Match 84.4%; Score 38; DB 4; Length 70;
Best Local Similarity 80.0%; Pred No. 22;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WVRWH 5
|:|:|
Db 4 WIRWH 8

RESULT 14
Q9SK65
ID Q9SK65 PRELIMINARY; PRT; 178 AA.
AC Q9SK65;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 20.3 kDa protein.
OS Macaca fascicularis (Orang Pating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MEDULLA OBLONGATA;
RA Osada N., Hida M., Kusuda T., Tanuma P., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
libraries."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB066522; BAB62201.1;
DR InterPro; IPR000372; LRR_Nterm.
DR Pfam; PF01462; LRRNT; 1.
KW Hypothetical protein.
SQ SEQUENCE 178 AA; 20242 MW; 4780043550D58F CDD64;

Query Match 84.4%; Score 38; DB 6; Length 178;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WVRWH 5
|:|:|
Db 4 WIRWH 8

RESULT 15

Q8WXJ6
ID Q8WXJ6 PRELIMINARY; PRT; 537 AA.
AC Q8WXJ6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Neurotrophin receptor tyrosine kinase type 2 (Fragment).
GN NTRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-2165693; PubMed-11798182;
RA Stojilov P., Castren E., Stamm S.;
RT "Analysis of the Human TrkR Gene Genomic Organization Reveals Novel
TrkR Isoforms, Unusual Gene Length, and Splicing Mechanism."
RL Biochem Biophys Res Commun 240:104-106(2002).
DR EMBL; AF410902; AAL67966.1;
DR InterPro; IPR000372; IPR_Nterm
DR Pfam; PF01462; LRRNT; 1.
DR SMART; SM00013; LRRNT; 1.
KW Receptor; Kinase.
FT NON_TER 537
SQ SEQUENCE 537 AA; 59166 MW; 5A8FA252A38710C1 CDD64;

Query Match 84.4%; Score 38; DB 4; Length 537;
Best Local Similarity 80.0%; Pred No. 16+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WVRWH 5
|:|:|
Db 4 WIRWH 8

Search completed: February 20, 2003, 10:49:18
Job time : 25.3333 secs

GenCore version 5.1.3
Copyright 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: February 20, 2003, 11-24-02 / Search time 12 seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09 912 414 9
Perfect score: 6
Sequence: 1 WXXWXP 6

Scoring table: OLIGOXY
Gapop 60.0, Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: PIP 73: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. NO is the number of results predicted by chance to have a
score greater than or equal to the score of the best result printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	21	2	PH1731
2	6	100.0	49	2	185 kDa glycoproph
3	6	100.0	59	2	hypothetical prote
4	6	100.0	61	2	hypothetical prote
5	6	100.0	65	2	light-harvesting p
6	6	100.0	71	2	hypothetical prote
7	6	100.0	73	2	hypothetical prote
8	6	100.0	76	2	hypothetical prote
9	6	100.0	79	2	hypothetical prote
10	6	100.0	79	2	hypothetical prote
11	6	100.0	83	2	hypothetical prote
12	6	100.0	86	2	conserved hypotet
13	6	100.0	87	2	prephage p12 prote
14	6	100.0	92	2	hypothetical 10.7K
15	6	100.0	92	2	hypothetical prote
16	6	100.0	93	2	hypothetical prote
17	6	100.0	97	2	hypothetical prote
18	6	100.0	99	2	hypothetical prote
19	6	100.0	101	2	conserved hypotet
20	6	100.0	103	2	conserved hypotet
21	6	100.0	105	2	hypothetical prote
22	6	100.0	106	2	hypothetical prote
23	6	100.0	106	2	hypothetical prote
24	6	100.0	106	2	hypothetical prote
25	6	100.0	109	2	hypothetical prote
26	6	100.0	109	2	hypothetical prote
27	6	100.0	109	2	hypothetical prote
28	6	100.0	109	2	hypothetical prote
29	6	100.0	109	2	hypothetical prote

30	6	100.0	110	2	AD0127	conserved hypotet
31	6	100.0	110	2	AG1523	hypothetical prote
32	6	100.0	116	2	P29966	13 heavy chain pre
33	6	100.0	116	2	383198	hypothetical prote
34	6	100.0	117	2	A95181	hypothetical prote
35	6	100.0	118	2	524527	13 heavy chain V r
36	6	100.0	119	2	524522	13 heavy chain V r
37	6	100.0	119	2	524492	13 heavy chain V r
38	6	100.0	119	2	524512	13 heavy chain V r
39	6	100.0	119	2	524493	13 heavy chain V r
40	6	100.0	119	2	524491	13 heavy chain V r
41	6	100.0	119	2	524525	13 heavy chain V r
42	6	100.0	119	2	524520	13 heavy chain V r
43	6	100.0	119	2	524519	13 heavy chain V r
44	6	100.0	119	2	524511	13 heavy chain V r
45	6	100.0	119	2	524494	13 heavy chain V r

ALIGNMENTS

RESULT 1

PH1731

Ig heavy chain V region (clone GCC-11) - mouse (fragment)

C/Species: Mus musculus (house mouse)

C/Date: 24 Feb-1994 #s-fragment-revision 04 Feb 1994 #text-change 10 Mar-1999

C/Accession: PH1731

F/McHeyzer Williams, M G, McLean, M J, Lajtor, F A, Rossal, G J V.

J Exp Med 178, 295-307, 1993

A/Title: Antigen-driven B cell differentiation in vivo.

A/Reference number: PH1675; MIMD:93301607; PMID:8315385

A/Accession: PH1731

A/Molecule type: mRNA

A/Residues: 1-21 <MOH>

A/Experimental source: B cell

A/Note: the authors translated the cDNA AOA for residue 13 as Ala

C/Superfamily: immunoglobulin V region, immunoglobulin V

C/Keywords: heterotetramer; immunoglobulin

Query Match

Best Local Similarity 100.0%; Score 6; DB 2; Length 21;

Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6

DB 16 WTSWSF 21

RESULT 2

I48681

185 kDa glycoprophosphoprotein - mouse (fragment)

C/Species: Mus musculus (house mouse)

C/Date: 02-Jul-1996 #sequence-revision 02-Jul 1996 #text-change 01 Jun 2002

C/Accession: I48681

R/White, M R; Hung, M C.

Oncogene 7, 677-683, 1992

A/Title: Cloning and characterization of the mouse neu promoter.

A/Reference number: I48681; MIMD:93304966; PMID:1348855

A/Accession: I48681

A/Status: preliminary; translated from GP/EMPI/PPRI

A/Molecule type: DNA

A/Residues: 1-49 <RES>

A/Accession: I48681; MIMD:93304966; PMID:1348855

C/Species: Mus musculus

C/Keywords: ATP; phosphoprotein

Query Match 100.0%; Score 6; DB 2; Length 49;

Best Local Similarity 50.0%; Pred. No: 1.4e+03;

Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6

DB 16 WTSWSF 21

A/Accession: AH2788
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-73 <KUR>
A/Cross-references: GR:AE009648, FIFN:AA142726.1, FID:J17741165, GCFPE:GN00186
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atul727
A/Map position: circular chromosome

Query Match 100.0%; Score 6; DB 2; Length 73;
Best Local Similarity 50.0%; Pred No 1 9e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 13 WNLWIF 18

RESULT 8
T14724
hypothetical protein Y1091 - Yersinia pestis plasmid pMT1
C/Species: Yersinia pestis
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Sep-1999
C/Accession: T14724; T15006
R/Hu, F, Piliot, J, McCready, P, Skowronski, E, Barnes, J, Kobayashi, A, Yamanaka, submitted to the EMBL Data Library, March 1998
A/Description: structural organization of virulence determinants in three Yersinia pestis
A/Reference number: Z18168
A/Accession: T14724
A/Status: preliminary; translated from GR/EMBL/DBEM
A/Molecule type: DNA
A/Residues: 1-76 <HUP>
A/Cross-references: EMBL:AF053947; MIM:30906296; FID:30906293, FIFN:AA019243.1
R/Lindler, L.E.; Plano, G.V.; Rutland, V.; Mayhew, G.F.; Plattner, P.P.
Infect. Immun. 66, 5731-5742, 1998
A/Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIMS plasmid
A/Reference number: Z18268; MIM:D199043898; PMID:9826348
A/Accession: T15006
A/Status: preliminary; translated from GR/EMBL/DBEM
A/Molecule type: DNA
A/Residues: 1-76 <LIR>
A/Cross-references: EMBL:AF074611, FID:31841093, FID:31841083, FIFN:AA089749.1
C/Genetics:
A/Gene: Y1091
A/Genome: plasmid pMT1

Query Match 100.0%; Score 6; DB 2; Length 76;
Best Local Similarity 50.0%; Pred No 1 9e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 50 WRTWNF 55

RESULT 9
S08449
hypothetical protein 9 - Spiroplasma virus 1
C/Species: Spiroplasma virus 1, Spv1
C/Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 07-Dec-1999
C/Accession: S08449
R/Pennaudin, J.; Aulio, F.; Vignault, J.C.; Bove, J.M.
Nucleic Acids Res. 18, 1293, 1990
A/Title: Complete nucleotide sequence of the genome of Spiroplasma citri virus spv1-PR82
A/Reference number: S08447; MIM:D190206709; PMID:2320423
A/Accession: S08449
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-78 <REN>
A/Cross-references: EMBL:X01344, MIM:J61093, FIFN:AA15727.1, FID:3008494
C/Genetics:
A/Genetic code: SGC3

Query Match 100.0%; Score 6; DB 2; Length 78;
Best Local Similarity 50.0%; Pred. No. 2e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 42 WNLWIF 47

RESULT 10
A64750
hypothetical protein b0249 - Escherichia coli (strain K12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C/Accession: A64750
R/Blattner, F.P.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Gill, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MIM:67436617; PMID:9278403
A/Accession: A64750
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-79 <BLAT>
A/Cross-references: GR:AE003177, GR:U00006, MIM:J199003, FIFN:AA077452.1, PMID:1286443;
A/Experimental source: strain K-12, substrain MG1655

Query Match 100.0%; Score 6; DB 2; Length 79;
Best Local Similarity 50.0%; Pred No 2e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 50 WPAWNP 55

RESULT 11
A97568
hypothetical protein AGR_C3173 [imported] - Agrobacterium tumefaciens (strain 789, 7100)
C/Species: Agrobacterium tumefaciens
C/Date: 20-Sep-2001 #sequence_revision 20-Sep-2001 #text_change 14-Jan-2002
C/Accession: A97568
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Gurolo, B.; Goldman, A.; Liu, P.; Wollam, C.; Allinger, M.; Dougherty, P.; Scott, C.; Lappas, C.; Markisz, B.; Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A/Reference number: A97359; PMID:11743194
A/Accession: A97568
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-83 <KUR>
A/Cross-references: GR:AE007863, FIFN:AA081408.1, FID:31517695, GCFPE:GN00169
C/Genetics:
A/Gene: AGR_C3173
A/Map position: circular chromosome

Query Match 100.0%; Score 6; DB 2; Length 83;
Best Local Similarity 50.0%; Pred No 2 1e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 23 WNLWIF 28

RESULT 12
T42185
conserved hypothetical protein L7085 - Escherichia coli plasmid p0157
C/Species: Escherichia coli
C/Date: 03-Sep-1999 #sequence_revision 03-Sep-1999 #text_change 02-Dec-1999
C/Accession: T42185
R/Rutland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Plattner, F.R.

Nucleic Acids Res. 26, 4196-4204, 1998

A>Title: The complete DNA sequence of the 1998 H5N1 avian influenza virus of H. N. 1998

A/Reference number: 222068; M01D:98591744; PMID:9722640

A/Accession: T42185

A/Status: preliminary; Unpublished from GP/FP/1/99

A/Molecule type: DNA

A/Residues: 1-92 <K1E>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: RNA

A/Residues: 1-92 <K1E>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

DB 07 WRTWTF 82

RESULT 15

A/Title: The complete DNA sequence of the 1998 H5N1 avian influenza virus of H. N. 1998

A/Reference number: 222068; M01D:98591744; PMID:9722640

A/Accession: T42185

A/Status: preliminary; Unpublished from GP/FP/1/99

A/Molecule type: DNA

A/Residues: 1-92 <K1E>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: RNA

A/Residues: 1-92 <K1E>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

A/Description: Human (H5N1) cells, the major B cell subset in the peripheral blood

A/Accession: S37501

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-87 <STO>

A/Cross-references: EMBL:AF074413; FIRM:AF074413

GenCorp version 5.1.1
Copyright (c) 1991 2003 GenCorp Inc.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:20:17 / Search time 11.733 seconds
(without alignments)
21.958 Million cell updates/sec

Title: us-09-912 414 9
Perfect score: 6
Sequence: 1 WXXWXP 6

Scoring table: OLIGODX
Gapop 60.0, Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post processing: listing first 45 summaries

Database: SwissProt_40.1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	6	100.0	65	LHA2_ECTHL	P46989 saccharomy
2	6	100.0	78	VG9_SPVIR	P71663 myobacteri
3	6	100.0	79	YKPF_ECORT	P74625 escherichia
4	6	100.0	84	ACRA_CROPS	P54348 crocidura r
5	6	100.0	84	ACRA_FELTA	P54350 felis silve
6	6	100.0	90	YKPF_ECORT	P58095 escherichia
7	6	100.0	101	YKPF_ECORT	P75147 mycoplasma
8	6	100.0	105	YKPF_ECORT	P38178 homo sapien
9	6	100.0	106	YKPF_ECORT	P38178 homo sapien
10	6	100.0	108	YKPF_ECORT	P38178 homo sapien
11	6	100.0	116	YKPF_ECORT	P19181 carassius a
12	6	100.0	128	YKPF_ECORT	P08656 staphylococ
13	6	100.0	129	YKPF_ECORT	P04107 homo sapien
14	6	100.0	129	YKPF_ECORT	P38857 saccharomyc
15	6	100.0	138	YKPF_ECORT	P75602 mycoplasma
16	6	100.0	139	YKPF_ECORT	P06784 mus musculu
17	6	100.0	144	YKPF_ECORT	P09492 aeropyrum p
18	6	100.0	145	YKPF_ECORT	P53311 saccharomyc
19	6	100.0	145	YKPF_ECORT	P53311 saccharomyc
20	6	100.0	147	YKPF_ECORT	P57085 methanococc
21	6	100.0	155	YKPF_ECORT	P08842 escherichia
22	6	100.0	157	YKPF_ECORT	P75147 mycoplasma
23	6	100.0	164	YKPF_ECORT	P00804 escherichia
24	6	100.0	165	YKPF_ECORT	P56678 vibrio algi
25	6	100.0	165	YKPF_ECORT	P13614 enterobacter
26	6	100.0	165	YKPF_ECORT	P53329 rhodobacter
27	6	100.0	166	YKPF_ECORT	P20209 sulfobobus
28	6	100.0	170	YKPF_ECORT	P17042 pseudomonas
29	6	100.0	171	YKPF_ECORT	P47657 mycoplasma
30	6	100.0	172	YKPF_ECORT	P75566 mycoplasma
31	6	100.0	176	YKPF_ECORT	P26701 bacterioph
32	6	100.0	177	YKPF_ECORT	P04107 homo sapien
33	6	100.0	182	YKPF_ECORT	P04107 bacterioph

34	6	100.0	196	YKPF_ECORT	P46989 saccharomy
35	6	100.0	200	YKPF_ECORT	P71663 myobacteri
36	6	100.0	208	YKPF_ECORT	P74625 escherichia
37	6	100.0	208	YKPF_ECORT	P54348 crocidura r
38	6	100.0	210	YKPF_ECORT	P54350 felis silve
39	6	100.0	213	YKPF_ECORT	P58095 escherichia
40	6	100.0	215	YKPF_ECORT	P75147 mycoplasma
41	6	100.0	217	YKPF_ECORT	P38178 homo sapien
42	6	100.0	217	YKPF_ECORT	P38178 homo sapien
43	6	100.0	217	YKPF_ECORT	P19181 carassius a
44	6	100.0	219	YKPF_ECORT	P08656 staphylococ
45	6	100.0	220	YKPF_ECORT	P04107 homo sapien

ALIGNMENTS

RESULT 1

LHA2_ECTHL STANDARD; PRT; 65 AA.

AC P80103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Light-harvesting protein B800/B850/1020, alpha-2 chain (LHA2-alpha 2)
DE (Antenna pigment protein, alpha-2 chain)
OS Escherichiodospira halochloris.
CC Bacteria, Proteobacteria, gamma subdivision, Escherichiodospiraceae;
CC Halorhodospira
OX NCBI_TaxID=1052;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1059;
PX METLIN=92349336; PubMed=1577009;
RA Wagner-Huber R, Brunisholz P A, Bissig I, Frank G, Suter F, Zuber H;
RA Eur. J. Biochem. 205:917-925(1992).

RT "The primary structure of the antenna polypeptides of Escherichiodospira halochloris and Escherichiodospira halophila. Four core-type antenna polypeptides in E. halochloris and E. halophila.";
RT Eur. J. Biochem. 205:917-925(1992).
CC FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS. SUBUNIT. THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE REACTION CENTER. THE NON FLUORESCENT GAMMA CHAINS MAY CONNECT THE ADDITIONAL COMPONENTS.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane. PIP, S23164; S23164
CC Interpro: IPR002361; AntennaA.
CC Interpro: IPR000066; Antenna_a/b.
CC Pfam: PF00556; LHC; 1
CC PROSITE: PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; light-harvesting polypeptide; transmembrane; Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1
FT TRANSMEM 14 34
FT DOMAIN 35 65
FT METAL 29 29
FT AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL.
FT MAGNESIUM (POTENTIAL).
SQ SEQUENCE 65 AA; 7688 MW; 55A4C306748E3D9A CRC64;

Query Match 100.0%; Score 6; DB 1; Length 65;
Best local similarity 60.0%; Prid No 7 9e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXXWXP 6

Db 2 WPLWKP 7

RESULT 2
VG9_SPVIR

SD SEQUENCE 84 AA; 6833 MW; 210072FP09670A6R CPO64;

Query Match 100.0%; Score 6; P1; Length 84;
Best Local Similarity 50.0%; Pred No 9.5e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWVF 6
Db 63 WKHWVF 68

RESULT 5

ACHA_FELCA STANDARD; PRT; 84 AA.
AC P54250;
DT 01-OCT-1996 (Rel 34, Created)
DT 01-OCT-1996 (Rel 34, Last sequence update)
DT 15-JUN-2002 (Rel 41, Last annotation update)
DE Acetylcholine receptor protein, alpha chain (Fragment)
GN CHRNA1.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Carnivora; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Felidae; Felis.
OX NCBI_Taxid=9685;
RN [1]
RP SEQUENCE FROM N A
RC TISSUE=Muscle;
RX MEDLINE=94345001; PubMed-7619217;
RA Barchan P, Owalla M, Fuchs S;
RT "The binding site of the nicotinic acetylcholine receptor in animal species resistant to alpha-bungarotoxin";
RL Biochemistry 34:9172-9176(1995).
CC "FUNCTION AFTER FINDING ACETYLCHOLINE, THE ACHR RESPONDERS BY AN EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA MEMBRANE."
CC -1- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA, DELTA, AND GAMMA (IN IMMATURE MUSCLE) & EPSILON (IN MATURE MUSCLE) CHAINS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC SIMILARITY BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U17007; AAA89114.1; -
DR InterPro; IPR001175; Neur_chan.
DR Pfam; PF02931; Neur_chan_LBD; 1.
DR PROSITE; PS00236; NEUROTF_ICH_CHANNEL; 1.
KW Receptor; Postsynaptic membrane; Ion channel; Glycoprotein;
KW Transmembrane.
FT NON_TER 1 1
FT DISULFID 7 21 BY SIMILARITY.
FT DISULFID 71 72 ASSOCIATED WITH RECEPTOR ACTIVATION (BY SIMILARITY).
FT CARBOHYD 20 20 N-LINKED (GLYCAN) (POTENTIAL).
FT NON_TER 84 84
SQ SEQUENCE 84 AA; 9423 MW; 596F9A10BA1P097F CPO64;

Query Match 100.0%; Score 6; P1; Length 84;
Best Local Similarity 50.0%; Pred No 9.5e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWVF 6
Db 63 WKHWVF 68

RESULT 6

YPJ1_ECOLI STANDARD; PRT; 90 AA.
ID YPJ1_ECOLI
AC P58025;
DT 16-OCT-2001 (Rel 40, Created)
DT 16-OCT-2001 (Rel 40, Last sequence update)
DT 15-JUN-2002 (Rel 41, Last annotation update)
DE Hypothetical protein ypj1.
GN YPJ1 OR B2641.1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N A
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed-9278503;
RA Blatter F F, Plunkett G III, Bloch C A, Ferna M T, Rutland V.,
PA Riley M., Collado-Vides J., Glasner J D., Rode C K., Mayhew G.F.,
RA Gregg J., Davis N W, Kirkpatrick H A, Gordon M A, Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K12";
RL Science 277:1233-1238(1997).
RN [2]
RP IDENTIFICATION, AND CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-1999).
CC "CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT HAD TO BE CORRECTED IN POSITION 86 AND TWO IN FRAME STOP CODONS HAD TO BE SUPPRESSED IN POSITIONS 8 AND 11."
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
IF EMBL; AE003449; .; NOT ANNOTATED CDS
DR EcoGene; EG13309; ypj1.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 90 AA; 10070 MW; E17EEF072E9E497 CPO64;

Query Match 100.0%; Score 6; P1; Length 90;
Best Local Similarity 50.0%; Pred No. 1e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWVF 6
Db 61 WPAWVF 66

RESULT 7

YG50_MYCPN STANDARD; PRT; 101 AA.
ID YG50_MYCPN
AC P75147;
DT 16-OCT-2001 (Rel 40, Created)
DT 16-OCT-2001 (Rel 40, Last sequence update)
DT 16-OCT-2001 (Rel 40, Last annotation update)
DE Hypothetical lipoprotein MPN650 precursor (E09_01101).
GN MPN650 OR MP192.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_Taxid=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plogens H., Plökl E., Li B.-C.,
RA Heilmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

RD Model: SwissProt Entry: Q14407 (AAB00968)
 RD SUBMITTER LOCATION: Attribution to the author by a lipid author
 CC (Potential).
 CC - 1 SIMILARITY: BELONGS TO THE MADS / MAD4 FAMILY.
 CC
 CC This SWISS PROT entry is copyrighted. It is protected through the copyright between the Swiss Institute of Bioinformatics and the EMBL institution. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as the statement is made that it was modified and this statement is not removed or changed by third parties without the explicit written permission of the license agreement from EMBL/Wellcome/SIB/ChEMBL or send an email to licensed@ebi.ac.uk.
 CC -----
 DR EMBL; AB000000, AAB00040.1,
 DR InterPro; IPRO01595; Lipoprotein_3.
 DR Pfam; PF00038; Lipoprotein_3_1.
 DR ProDom; PD0000276; Lipoprotein_3_1.
 DR ProSite; PS00003; PROSITE_LipoProtein_3.
 KM Hypothetical Protein, HAP10000, Neisseria meningitidis, complete proteome.
 KW SIGNAL: 1 POTENTIAL.
 FT CHAIN: 26
 FT LIPID: 26
 FT REFERENCE: 101 AA; 1100 MW; C71F0B9EEFF1975735 CD0564;
 SC

Query Match	100.0%	Score 67	DB 1	Length 101
Similarity	50.0%	Prod NC_011416.03		
Matches	3	Conserved	0	Info-Base

QY	1	XXXXXXF	6
QB	95	WINNCKF	100

[illegible][illegible]

```
CC .....  
DR EMBL; AF060697; AMO5428.1; ..  
DE EMBL; AF060697; AAO32450.1; ..  
ID EMBL; M000149; AA01168.1; ..  
TS CDSW, NCBI, EMBL, REFSEQ.  
DK MIM; 603848; ..  
EW Oxidized; d; d; d; MAD; Ubiquitin; Mitochondrial; Tryptic; pep+ide;  
FI TRANSIT; ..; 3; MITOCHONDRION (BT SIMILARITY);  
FI CHAIN; 24; 100; NADH DEHYDROGENASE SUBUNIT  
FT SUBUNIT.  
SC SEQUENCE; 100 AA; 1000 NM; 50000 RESIDUES; OKCD4;
```

Query Match	100.0%	Score 67	DB 17	Length 105
Host Local Similarity	10.0%	Fold_M: 1.1e-03		
Matches	3	Conservative	0	Indels 0
		Mismatches	0	Gaps 0

OY	1	WXXMXF	6
		:	:
		:	:
		:	:
Db	68	WILWRF	73

[illegible]


```

00 (MIRA)
00 - SUBCELLULAR LOCATION: integral membrane protein.
00 -----
00 This entry represents a highly conserved protein found
00 between the Swiss Institute of Bioinformatics and the EMBL database.
00 The European Bioinformatics Institute. There are no restrictions on its
00 use by non-profit institutions as long as its name is in no way
00 modified and this statement is not removed. Tags by and for commercial
00 entities require prior agreement from http://www.ebi.ac.uk/submit/.
00 or send an email to license@ebi.ac.uk.
00 -----
00 DR EMBL; L29436; AAA98244.1; -.
00 DR FRL; D25504; L25504.
00 KW Transport; Mercury; resistance; Transmembrane; Mercury; Plasmid.
00 FT TRANSMEM 28
00 FT TRANSXEN 71 POTENTIAL.
00 FT METAL 47 31 POTENTIAL.
00 FT METAL 47 47 HG(2+) (POTENTIAL).
00 FT METAL 48 48 HG(2+) (POTENTIAL).
00 FT METAL 114 114 HG(2+) (POTENTIAL).
00 FT METAL 115 115 HG(2+) (POTENTIAL).
00 SC SEQUENCE 128 AA; 14182 MW; C9974E2B4A8AF91 CRC64;
00 -----
00 Query Match Score 57 Pz 17 Length 1287
00 Best Local Similarity 50.0%, Prod Ref 130433,
00 Matches 37 Of 128710, 3, Mismatched, 1, Indel, 1, Gap
00 -----
00 CY 1 MAXMAXF 6
00 Db 28 MGWMAF 23
00 -----
00 RESULT 23
00 KV3H_HCVAN STANDARD, FEI, 123 AA
00 AC P04207;
00 DT 20-MAR-1987 (Rel. 04, Created)
00 DT 01-NOV-1998 (Rel. 16, Last sequence update)
00 DT 15-JUN-1999 (Rel. 28, Last annotation update)
00 DE IG Kappa Chain V-II Region CLL Proliferation (Human);
00 OS Homo sapiens (Human).
00 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
00 CC Mammalia; Eutharia; Primates; Catarrhini; Hominidae; Homo.
00 CX NCBI_TaxID:9606;
00 RN SEQUENCE FROM N.A.
00 RP MEDLINE=86177570; PubMed=1083417;
00 RA Girib F.E., S-Gut T., F.J.S., Hollmann J.G., Call J.S., Chen F.T.,
00 RA Golden R., Carson D.A.;
00 RA "Cloning and sequence determination of a human lymphoid factor
00 RT "IgG chain gene."
00 RL Proc Natl Acad Sci USA 93:11511-11515(1996)
00 CC -----
00 This entry represents a highly conserved protein found
00 between the Swiss Institute of Bioinformatics and the EMBL database.
00 The European Bioinformatics Institute. There are no restrictions on its
00 use by non-profit institutions as long as its name is in no way
00 modified and this statement is not removed. Tags by and for commercial
00 entities require prior agreement from http://www.ebi.ac.uk/submit/.
00 or send an email to license@ebi.ac.uk.
00 -----
00 DR EMBL; M12740; AAA58992.1;
00 DR PIR; A01898, K3HUCL.
00 DR HGSP; P80362, IWTU.
00 DR InterPro; IPR003000, Ig_MHC.
00 DR InterPro; IPR003006, Ig_V.
00 DR Pfam; PF00347, Ig_I.
00 DR SMART; SM00426, IGV_1.
00 KW Immunoglobulin V region; Signal.
00 FT SIGNAL 1
00 FT CHAIN 21 129 IG KAPPA CHAIN V-II REGION CLL.
00 FT DOMAIN 21 43 FRAMEWORK-1.
00 FT DOMAIN 44 64 COMPLEMENTARITY DETERMINING-1.
00 FT DOMAIN 65 69 FRAMEWORK-2.

```

```

01 DOMAIN 70 70 COMPLEMENTARITY DETERMINING 2.
02 DOMAIN 70 70 FRAMWORK-3.
03 DOMAIN 100 118 COMPLEMENTARITY DETERMINING-3.
04 DOMAIN 100 118 FRAMWORK-3.
05 DOMAIN 100 118 COMPLEMENTARITY DETERMINING-3.
06 DOMAIN 100 118 FRAMWORK-3.
07 NON TER 129 129 BI SIMILARITY.
08 SEQUENCE 100 AA, 14271 MW, 11228116600014 CRO64;
Query Match 100.0%; Score 6; DB 1; Length 129;
Best Local Similarity 50.0%; Pval NO 1.3e+03;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

01 1 WXXNKF 6
02 111111
03 114 WRPWTF 119
04 114 WRPWTF 119

RESULT 14
YHM2 YEAST
01 YHM2 YEAST STANDARD; PRT; 129 AA.
02 DT 01-FEB-1995 (Rel. 31, Created)
03 DT 01-FEB-1995 (Rel. 31, Last sequence update)
04 DT 16-OCT-2001 (Rel. 40, Last annotation update)
05 HYPERLINKED TO A DNA REGION IN FLUENT FILE: ALE130002 region.
06 YHR162W.
07 Saccharomyces cerevisiae (Baker's yeast).
08 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
09 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
10 NCBI TaxID=4932;
11 [1]
12 SEQUENCE FROM N.A.
13 STRAIN=S288C / AB972;
14 MEDLINE 747844; PubMed 841147;
15 Johnston M, Andrews S, Binkman R, Cooper D, Ding H, Dover D,
16 Du Z, Favell A, Fulton L, Gattung S, Geisel C, Kirsten D,
17 Korda T, Miller L, O'Brien J, Schmeichler J, Jackson Y,
18 Lathrop D, Louis E, MacI C, Mardin E, Marzke S, Mouser T,
19 Mow W, Pridmore D, Rieger J, Peter H, Tronicko F, Vaughan K,
20 Vignati D, Wilcox L, Wolfrum F, Wollstein R, Wilson R,
21 Vaudin M;
22 "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
23 VIII."
24 Science 265:1207-1209 (1994).
25 -1- SIMILARITY: BELONGS TO THE UPT0041 (U-44) FAMILY.
26 -----
27 This SWISS-PROT entry is a duplicate of the UPT0041 entry. A collaboration
28 between the Swiss Institute of Bioinformatics and the EMBL database on the
29 the European Bioinformatics Institute. There are no restrictions on its
30 use by any party. The content is in no way
31 modified and this statement is not removed. Usage by any party (commercial
32 entities requires a license agreement. See http://www.ebi.ac.uk/announcements/
33 for more details.
34 -----
35 EMBL: 000027; AAB69009.1;
36 EIR, 34302, 31502.
37 SGR, 0001005; YHR162W.
38 Interim, 1141111, UPT0041.
39 Pfam, Pfam0041; 1.
40 KW Hypothetical protein.
41 SEQUENCE 100 AA, 14271 MW, 11228116600014 CRO64;
Query Match 100.0%; Score 6; DB 1; Length 129;
Best Local Similarity 50.0%; Pval NO 1.3e+03;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

01 1 WXXNKF 6
02 111111
03 114 WRPWTF 119
04 114 WRPWTF 119

RESULT 15
YAG1 MYCFN

```

```

ID      YA91 MYCPN      STANDARD;      PRT;      138 AA.
AC      P75602;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein MPN091 (R02_orf138).
GN      MPN091 OR MP064.
OS      Mycoplasma pneumoniae.
OC      Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX      NCBI_TaxID=2104;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 29342 / M129;
RX      MEDLINE=97105865; PubMed=8948633;
RA      Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA      Herrmann R.;
RT      "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT      pneumoniae.";
RL      Nucleic Acids Res. 24:4420-4449(1996).
CC      -1- SIMILARITY: ALMOST IDENTICAL TO M.FNECMONIAAE MPN413 AND MPN463.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE000008; AAB95711.1; -.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 138 AA; 15763 MW; F52380B428F42709 CRC64;

Query Match      100.0%; Score 6; DB 1; Length 138;
Best Local Similarity 50.0%; Pred. No. 1.4e+03;
Matches 3; Conservative 3; Mismatch 0; Indels 0; Gaps 0;

QY      1 WXXWXF 6
       |::||
Db      36 MSGWSF 41
    
```

Search completed: February 20, 2003, 11:24:57
 Job time : 12.3333 secs

GenCore version 5.1.3
Copyright (c) 1992 2003 GenCyber Inc.

OM protein: protein search, using sw model

Run on: February 20, 2003, 11:33:27, Search time 46 Secs (without alignments)
26,876 Million cell updates/sec

Title: US-09-912-414-9

Perfect score: 6

Sequence: 1 WXXWXP 6

Scoring table: Gapped 60 0, Gapext 60 0

Searched: 67580 seqs, 26634715 residues

Word size: 0

Total number of hits satisfying all search parameters: 67580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPRMBL_21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriapi:
17: sp_archaeapi:

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	28	6 Q62821	Q62821 bubalus bub
2	6	100.0	31	8 Q94NH2	Q94NH2 titratorin s
3	6	100.0	36	15 Q905D7	Q905D7 human immun
4	6	100.0	48	11 Q63165	Q63165 rattus norv
5	6	100.0	49	11 Q08615	Q08615 mus musculu
6	6	100.0	50	4 Q9NWT0	Q9NWT0 homo sapien
7	6	100.0	59	16 Q9PH20	Q9PH20 xylella fas
8	6	100.0	61	16 Q96619	Q96619 escherichia
9	6	100.0	64	4 Q8WYM9	Q8WYM9 homo sapien
10	6	100.0	70	2 Q9EXE4	Q9EXE4 lactobacill
11	6	100.0	71	17 Q9HST8	Q9HST8 halobacteri
12	6	100.0	74	2 Q50168	Q50168 mycobacteri
13	6	100.0	76	16 Q96793	Q96793 yersinia pe
14	6	100.0	77	2 Q9S4W3	Q9S4W3 escherichia
15	6	100.0	79	2 Q9WTF7	Q9WTF7 escherichia
16	6	100.0	79	16 Q9JMR3	Q9JMR3 escherichia

17	6	100.0	80	16 Q93GP6	Q93GP6 salmonella
18	6	100.0	83	16 Q8UEM9	Q8UEM9 agrobacteri
19	6	100.0	85	2 Q9FET3	Q9FET3 escherichia
20	6	100.0	86	2 Q923P9	Q923P9 escherichia
21	6	100.0	86	2 Q9N1M2	Q9N1M2 solanum cha
22	6	100.0	87	18 Q905E4	Q905E4 lactococcus
23	6	100.0	89	10 Q8W1L9	Q8W1L9 solanum pin
24	6	100.0	89	10 Q8W1L9	Q8W1L9 solanum pin
25	6	100.0	92	10 Q8W1M0	Q8W1M0 solanum bul
26	6	100.0	92	12 Q9NWT0	Q9NWT0 xylella fas
27	6	100.0	93	10 Q8W1M1	Q8W1M1 solanum bul
28	6	100.0	93	16 Q905E4	Q905E4 lactococcus
29	6	100.0	94	4 Q96793	Q96793 yersinia pe
30	6	100.0	96	4 Q96793	Q96793 yersinia pe
31	6	100.0	97	16 Q9K5V6	Q9K5V6 bacillus ha
32	6	100.0	99	10 Q94NH2	Q94NH2 titratorin s
33	6	100.0	99	16 Q96793	Q96793 yersinia pe
34	6	100.0	101	5 Q96793	Q96793 yersinia pe
35	6	100.0	103	10 Q9X052	Q9X052 titratorin s
36	6	100.0	103	16 Q9K5V6	Q9K5V6 bacillus ha
37	6	100.0	103	16 Q9X052	Q9X052 titratorin s
38	6	100.0	105	11 Q905D7	Q905D7 human immun
39	6	100.0	105	16 Q9X052	Q9X052 titratorin s
40	6	100.0	106	16 Q9PES0	Q9PES0 mus musculu
41	6	100.0	106	16 Q9X1V7	Q9X1V7 xylella fas
42	6	100.0	107	16 Q96793	Q96793 yersinia pe
43	6	100.0	107	2 Q96793	Q96793 yersinia pe
44	6	100.0	108	16 Q9YKX8	Q9YKX8 anabaena sp
45	6	100.0	108	16 Q96793	Q96793 yersinia pe

ALIGNMENTS

RESULT 1					
ID	Q62821	PRELIMINARY;	PRT;	28 AA.	
AC	Q62821;				
DT	01-AUG-1998 (TREMBLrel. 07, Created)				
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	Lysozyme (Fragment).				
OS	Bubalus bubalis (Domestic water buffalo).				
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Fumiantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bubalus.				
OX	NCBI_TaxID=89462;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
FC	TISSUE=BL00D;				
PA	Das P., Tiwari G., Garg I. C.;				
RT	"Characterisation of lysozyme encoding gene promoter in Bubalus				
RT	bubalis"; (MAR-1998) to the EMBL/GenBank/DBJ databases.				
PL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; A025012; CAA02342.1; .				
FT	NON TER	29	28		
SQ	SEQUENCE	28 AA;	3160 MW;	8806FCAD7DD829D CRC64;	

Copy Match 100.0%, Score 6, PP 6, Length 28;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

QY 1 WXXWXP 6
DB 10 WSWWTP 15

RESULT 2

ID Q94NH2 PRELIMINARY; PRT; 31 AA.
AC Q94NH2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

QY 1 WXXWXP 6
DB 3 WDRWRF 8

RESULT 9

Q9HST8 PRELIMINARY; PRT; 64 AA.
AC Q9HST8;
DI 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eulalia; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC Tissue-BRAIN;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/CCPDB databases.
KW Hypothetical protein
FT NON-TER 1
SQ SEQUENCE 64 AA, 1480 MW, 33722aa/75401aa/27064,
[11]
Query Match 100.0%; Score 6, Db 4, Length 64;
Best Local Similarity 50.0%; Pred. No. 3.4e+03;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

QY 1 WXXWXP 6
DB 16 WDRWRF 21

RESULT 10

Q9EXE4 PRELIMINARY; PRT; 70 AA.

AC Q9EXE4;
DI 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
OS Purative thymidylate synthase (Fragment).
OC Bacteria; Firmicutes; Bacilli; Clostridia; Clostridiales;
OC Lactobacillaceae; Lactobacillus.
OX NCBI_TaxID=1549;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN 154X3;
RA Basco A.; Campanile C.; Ferrara L.; Damiani G.;
RT "Sequencing and analysis of DNA fragments obtained by random amplified
polymorphic DNA (RAPD) technique from Lactobacillus sakei strains";
RL Submitted (JUN-2002) to the EMBL/GenBank/CCPDB databases.
DI EMBL; AJ404622; CAC00646.1;
DR HSSP; P03469; ITSY;
DR InterPro; IPR000099; Thymidylat_syn
DR Pfam; PF00303; Thymidylate synthase;
DR ProDom; PD001180; Thymidylat_syn; 1
FT NON-TER 1
FT NON-TER 70
SQ SEQUENCE 70 AA, 1483 MW, 4332aa/10741aa/2574,
[11]
Query Match 100.0%; Score 6, Db 4, Length 70;
Best Local Similarity 50.0%; Pred. No. 3.6e+03;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

QY 1 WXXWXP 6
DB 60 WDRWRF 65

RESULT 11

Q9HST8 PRELIMINARY; PRT; 71 AA.
AC Q9HST8;
DI 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
OS Vng0079h.
CN VNG0079H.
AC Halobacterium sp. (Strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=20504493; PubMed=11016950;
RA Ng M.Y., Kennedy S.P., Mahulis G.G., Bergquist B., Fan M.,
RA Shukla H.D., Lasky G.R., Baliga N.S., Thorson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall C., Dahl T.A., Wells R., Goo Y.A.,
RA Leithausen R., Kojima F., Croft P., Paton M.C., Hough D.W.,
RA Maddocks J.G., Jatkowski F.E., Krebs M.F., Angerino C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischer M., Spedlich C.H., Jung K.-H.,
RA Alam M., Freitas T., Hoi S., Daniels C.J., Dennis R.P., Omer A.D.,
RA Ebhardt H., Lowe T.W., Ling F., Riley M., Reid T., DiCarma C.;
RL Genome sequence of Halobacterium species NRC-1.";
DI Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DI EMBL; AF094907; A018714.1;
KW Complete proteome.
FT SEQUENCE 71 AA, 1413 MW, 4474aa/11023aa/60064,
[11]
Query Match 100.0%; Score 6, Db 17, Length 71;
Best Local Similarity 50.0%; Pred. No. 3.7e+03;
Matches 4, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

QY 1 WXXWXP 6
DB 49 WDRWRF 54

RESULT 12

Q90168 PRELIMINARY; PRT; 74 AA.
AC Q90168;
DI 01 NOV 1996 (TREMBLrel. 01, Created)
DT 01 NOV 1996 (TREMBLrel. 01, Last sequence update)
DE 01 NOV 1996 (TREMBLrel. 01, Last annotation update)
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN 1;
RP SEQUENCE FROM N.A.
RA Smith D.R.;
RL Submitted (NOV 1996) to the EMBL/GenBank/CCPDB databases.
DI EMBL; U01143;
DR Robinson R.;
DI EMBL; U01143; AAA6314.1;
SQ SEQUENCE 74 AA, 1343 MW, 3315aa/8317aa/3504,
[12]
Query Match 100.0%; Score 6, Db 4, Length 74;
Best Local Similarity 50.0%; Pred. No. 3.8e+03;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

QY 1 WXXWXP 6
DB 59 WDRWRF 64

RESULT 13

```

068793
ID 068793 PRELIMINARY; PRT; 76 AA.
AC 068793;
DT 01-AUG-1998 (TRENBLREL. 07, Created)
DT 01-AUG-1998 (TRENBLREL. 07, Last sequence update)
DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
DE Hypothetical protein Y1001;
GN Y1001 OR YPMT1.60C.
OS Yersinia pestis.
OG Plasmid PMT1 (PMT-1).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM5;
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCready P., Skowronski E., Barnes J.,
RA Kobayashi A., Rubaker R.R., Garcia E.;
RT "Structural organization of virulence-associated plasmids of Yersinia
RT pestis.";
RL J. Bacteriol. 180:5192-5202(1998)
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM10+;
RX MEDLINE=99043888; PubMed=9826348;
RA Lindler B.E., Plano G.V., Burland V., Mayhew G.F., Plattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT KIM5 plasmid encoding murine toxin and capsular antigen ";
RL Infect Immun 66:5731-5742(1998)
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Co 32 / PT/VAF ORIENTALIS; PLASMID-PMT1 (PMT 1);
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.P., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.I.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies P.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Jarvis S.A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skellern J., Stevens K., Whitehead S., Barrall P.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague ";
RL Nature 413:523-527(2001).
DR EMBL; AF053947; AAC13243.1; -
DR EMBL; AF074611; AAC82749.1; -
DR EMBL; AL117211; CAB55242.1; -
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 76 AA; 8807 MW; 7EF4937755036BB CRC64;

Query Match 100.0%; Score 6; DB 16; Length 76;
Best Local Similarity 50.0%; Pred. No. 3.9e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 50 WRTWNP 55

RESULT 14
O9S4W3 PRELIMINARY; PRT; 77 AA.
AC O9S4W3;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-MAY-2000 (TRENBLREL. 13, Last annotation update)
DE Hypothetical 9.1 kDa protein.
OS Escherichia coli.
OG Plasmid F.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

```

FX MEDLINE=99296678; PubMed 10366527;
FA Manwaring N.P., Skurray P.A., Fitch N.;
RT "Nucleotide sequence of the F plasmid leading region.";
RL Plasmid 41:219-225(1999).
DR EMBL; AF106323; AAD47188.1; -
FW Hypothetical protein; Plasmid.
SQ SEQUENCE 77 AA; 9283 MW; 36345FF02678677A CRC64;

Query Match 100.0%; Score 6; DB 2; Length 77;
Best Local Similarity 50.0%; Pred. No. 4e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 53 WRTWNP 58

RESULT 15
O9WTF7 PRELIMINARY; PRT; 79 AA.
AC O9WTF7;
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE YdaA protein.
GN YDEA.
OS Escherichia coli.
OG Plasmid p100.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Sampei G., Mizobuchi K.;
RT "Organization and diversification of plasmid p100, complete
RT nucleotide sequence of the p100 genome ";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBP database.
DR EMBL; AF000340; PAA78827.1; -
SQ SEQUENCE 79 AA; 9208 MW; 7177949F777767 F764;

Query Match 100.0%; Score 6; DB 2; Length 79;
Best Local Similarity 50.0%; Pred. No. 4e+03;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 55 WRTWNP 60

```

Search completed: February 20, 2003, 11:28:57
 Job time : 47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using fw model

Run on: February 20, 2003, 11:21:07, Search time 29.333 Seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912 414 9

Perfect score: 6

Sequence: 1 WYXXXP

Scoring table:

OLIGOX
Gapop 60.0, Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size: 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum IP seq length: 200000000

Post-processing: listing first 45 summaries

Database: A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	6	15	AA57386 Peptide for treati
2	6	100.0	6	15	AA57391 Peptide for treati
3	6	100.0	6	21	AA57392 Peptide which bind
4	6	100.0	6	21	AA57397 Peptide which bind
5	6	100.0	6	21	AA57499 Peptide which bind
6	6	100.0	6	21	AA57505 Peptide which bind
7	6	100.0	6	21	AA57506 Peptide which bind
8	6	100.0	6	21	AA57508 Peptide which bind
9	6	100.0	7	22	AA575181 Polyglutamine repe
10	6	100.0	7	22	AA575777 H11 binding site c

11	6	100.0	8	15	AA57399 Peptide for treati
12	6	100.0	9	21	AA57498 Peptide which bind
13	6	100.0	9	22	AA573672 Peptide for treati
14	6	100.0	10	15	AA57400 Peptide for treati
15	6	100.0	11	19	AA572149 Haemophilus influe
16	6	100.0	11	21	AA570714 Peptide for treati
17	6	100.0	12	15	AA57401 Peptide for treati
18	6	100.0	12	21	AA574104 Oregon green 514 b
19	6	100.0	12	21	AA574154 Dystrophin WW doma
20	6	100.0	13	18	AA5738112 Human Elys binding
21	6	100.0	13	23	AB574201 Human HLA-DP restr
22	6	100.0	14	22	AA570760 Eptope derived fr
23	6	100.0	15	20	AA5730351 Streptococcus pneu
24	6	100.0	15	21	AA576504 Human ATP-dependen
25	6	100.0	15	23	AA577026 Streptococcus pneu
26	6	100.0	15	23	AA576245 Peptide which bind
27	6	100.0	15	23	AA574924 Insulin/insulin-11
28	6	100.0	17	21	AA5709193 Hepatitis B virus
29	6	100.0	19	22	AA576504 Novel human digno
30	6	100.0	20	23	AA570618 Consensus signal p
31	6	100.0	20	23	AA570940 Consensus signal p
32	6	100.0	20	23	AA570772 Consensus signal p
33	6	100.0	21	22	AB574111 Consensus signal p
34	6	100.0	21	22	AB5728607 Consensus signal s
35	6	100.0	22	22	AB579009 Consensus signal p
36	6	100.0	22	22	AA573402 Consensus signal p
37	6	100.0	22	22	AA572570 Consensus signal p
38	6	100.0	22	22	AA573132 Consensus signal p
39	6	100.0	22	22	AA572406 Consensus signal p
40	6	100.0	22	23	AA579514 Consensus signal p
41	6	100.0	22	23	AA573324 Consensus signal p
42	6	100.0	22	23	AA572951 Consensus signal p
43	6	100.0	22	23	AA573154 Consensus signal p
44	6	100.0	22	23	AA572223 Consensus signal p
45	6	100.0	22	23	AA575412 Tumor necrosis fa

ALIGNMENTS

RESULT 1
AA57386
ID AA57386 standard, peptide: 6 AA.
XX
AC AA57386;
XX
XX 21-MAR-1995 (first entry)
DT
XX
XX Peptide for treating diseases related to anti FNA antibodies.
DE
XX Carrier; absorbing agent, treatment; anti-FNA antibody, immune complex.
KW
XX
XX Synthetic
OS
XX
XX JP66192290-A.
XX
XX 12-JUL-1994.
PD
XX
XX 18-JAN-1993; 93JP-0006098.
PF
XX
XX 30-SEP-1992; 93JP-0261821.
PR
XX
XX (KUPRS) KUPAPAY CO LTD
PA
XX
XX WPI, 1994-260510/32.
DR
XX
XX A peptide and an adsorbing agent prep. by immobilising it on a
PT carrier - useful for treatment of diseases related to anti DNA
PT antibodies and immune complexes
XX
XX Disclosure; Page 11; 14pp; Japanese.
PS
XX
XX The sequences given in AA57386-413 are peptides which are all covered

```

CC by the standard generic formula.
CC      H X (A B n Y Z
CC      A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC      B = Trp, Phe, Asn or Glu;
CC      X and Y - a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC      consisting of 2-10 residues, provided that at least one of
CC      X or Y are present;
CC      Z = CH or NH2; and
CC      n = 2-5.
CC      These peptides may be immobilised on a carrier in the preparation of an
CC      absorbing agent which may be used in the treatment of diseases related
CC      to anti-DNA antibodies and/or immune complex.
CC      XX
SC      Sequence      6 AA;
      Query Match      100.0%      Score 67      DB 15      Length 67
      Best Match Similarity: 100.0%      E-Val: 1e-11
      Matches: 3, Conservation: 3, Mismatches: 0, Gaps: 0,
      07      1  WYKWNFF 6
      1111111
      Db      1  WYFWWF 6

```

```

XX RESULT 2
XX AAR57391
XX ID AAR57391 standard; Protein; 6 AA.
XX AC AAR57391
XX DT 21-MAR-1995 (first entry)
XX DE Peptide for treating diseases related to anti-DNA antibodies.
XX KW Carrier, absorbing agent, treatment, anti-DNA antibody, immune system
XX OS Synthetic
XX ST 01032000 A.
XX PD 12-JUN-1994.
XX PF 18-JAN-1993, 000P-000000008.
XX PR 30 DEC 1992, 000P-000000000.
XX PA (KURS ) KUPAPAY ON RTP
XX NT, 1994 000P-000000000.
XX LT
XX PT A peptide and an absorbing agent. By immobilizing it on a
XX PT carrier - useful for treatment of diseases related to anti-DNA
XX PT antibodies and immune complexes
XX PS
XX PS Disclosure; Page 11; 14pp; Japanese
XX CC The sequences given in AAR57396-419 are part of which are all covered
XX CC by the claimed generic formula:
XX CC H-X-(A-B)n-Y-Z
XX CC A - Trp, Phe or a peptide fragment consisting of 2 residues;
XX CC B = Trp, Phe, Asn or Glu;
XX CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
XX CC consisting of 2-10 residues, provided that at least one of
XX CC X or Y are present;
XX CC Z = OH or NH2; and
XX CC n = 2-5.
XX CC These peptides may be immobilized on a carrier in the preparation of an
XX CC absorbing agent which may be used in the treatment of diseases related
XX CC to anti DNA antibodies and/or immune complex.
XX CC
XX Sequence 6 AA;
SQ

```

POST TEST: Similarity: 10.0%,	Pred. No. 7.80(05)
Matches: 2,	Conservative: 0,
	Mismatches: 0,
	Indels: 0,
	Gaps: 0,

[illegible]

KW DNA binding, transcription factor; E2F, E2F-1; cell cycle; DP-1;
KM activation; transcription, apoptosis; proliferative disorder;
XX psoriasis; restenosis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2 /note= "Any amino acid"
FT Misc-difference 3 /note= "Any amino acid"
FT Misc-difference 5 /note= "Any amino acid"
FT Misc-difference 5 /note= "Any amino acid"
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PP 26 JAN 2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Claim 4; Page 9; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 100.0%; Pred No 7 Re+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
DB 1 WXXWXP 6

RESULT 5
AAB01499
ID AAB01499 standard; peptide; 6 AA.
XX
AC AAB01499;
XX
DT 0A-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding; transcription factor; E2F, E2F-1; cell cycle; DP-1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2 /note= "Any amino acid"
FT Misc-difference 3 /note= "Any amino acid"

FT /note= "Any amino acid"
XX
PN WO200044771-A1.
XX
PD 03 AUG 2000.
XX
PF 26 JAN 2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Claim 4; Page 9; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 83.3%; Pred No 7 Re+05;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
DB 1 WXXWHP 6

RESULT 6
AAB01505
ID AAB01505 standard; peptide; 6 AA.
XX
AC AAB01505;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding; transcription factor; E2F, E2F-1; cell cycle; DP-1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN 2000; 2000WO-GB00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DR WPI; 2000-532806/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of

PT cancer
XX
PS Example: Page 26; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 100.0%; Prod No. 7 0a10f;
Matches 2; Conservation 0; Mismatch 0; Indels 0; Gaps 0;
QY 1 WXXWXP 6
DB 1 WAWWHP 6

RESULT 7
AAB01506 (standard; peptide; 6 AA)
XX
AC AAF01506;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW DNA binding, transcription factor; E2F; E2F 1; cell cycle; DP 1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WC2000044771-A1.
XX
PD 03-AUG-2000
XX
PF 26-JAN-2000; 2000WO GP00227.
XX
PR 26-JAN 1999; 993P 0001710
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Korteemann RB, Moutglund S;
XX
DR WRI; 2000-032897/49
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example, Page 26; 42pp; English
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP 1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 100.0%; Prod No. 7 0a10f;
Matches 2; Conservation 0; Mismatch 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
DB 1 WAWWHP 6

RESULT 8
AAB01508
XX
AC AAB01508; standard; peptide; 6 AA.
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding, transcription factor; E2F; E2F 1; cell cycle; DP-1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WC2000044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN 2000; 2000WO GP00227.
XX
PR 26 JAN 1999; 993P 0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Korteemann RB, Moutglund S;
XX
DR WRI; 2000-032897/49.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example; Page 26; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F 1/DP 1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 100.0%; Prod No. 7 0a10f;
Matches 2; Conservation 0; Mismatch 0; Indels 0; Gaps 0;
QY 1 WXXWXP 6
DB 1 WAWWHP 6

RESULT 9
AAB01511
XX
AC AAB01511; standard; peptide; 7 AA.
XX
DT 18-DEC-2001 (first entry)
XX
DE Polyglutamine repeat domain binding peptide, QPB1-M7PD.
XX
KW Polyglutamine repeat domain, Huntington, ataxin,
KW Huntington's disease, dentatorubral pallidum degeneration; DRP1A;
KW

KW spinobulbar muscular atrophy; spinocerebellar ataxia; neurotic
XX Unidentified.
OS
XX WO200168678-A2.
PN
XX 20-SEP-2001.
PD
XX 14-MAR-2001; 2001WO-0508222.
PF
XX 16-MAR-2000; 2000US 189781P.
PR
XX (UYDU) UNIV DUKE.
PA
XX Strittmatter WJ, Burke JP, Nagai Y;
PI
XX WPI; 2001 616286/71
DR
XX
PT New polypeptides which selectively bind to expanded polyglutamine
PT repeat domains, useful for treating neurodegenerative diseases, e.g.
PT Huntington's disease
XX
XX Example 13; Page 14; 59pp; English.
PS
XX The present sequence is a polyglutamine repeat domain binding peptide.
CC The peptide is useful for treating a cell that contains and expresses a
CC protein having an expanded polyglutamine region (e.g. huntingtin;
CC atrophin 1, ataxin 1, 2, 6 or 7; or androgen receptor protein), or
CC treating a neurodegenerative disease characterised by the presence of
CC expanded polyglutamine repeats, e.g. Huntington's disease, dentatorubral
CC pallidolysian atrophy (DRPLA), spinobulbar muscular atrophy, and
CC spinocerebellar ataxia types 1, 2, 3, 6 and 7
XX
SQ Sequence 7 AA;

Query Match 100.0%, Score 6, DB 22; Length 7;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 1 WKWWPF 6

RESULT 10
AAM45777
ID AAM45777 standard; Peptide; 7 AA.
XX
AC AAM45777;
XX
DT 25-OCT-2001 (first entry)
XX
DE H11 binding site consensus conforming peptide (CCP) #2048.
XX
KW Antigen-binding, T-cell, diagnosis, stress protein-peptide complex, effect,
KW immunogenically cross-reactive; cancer; immunogenic cancer cell;
KW cytostatic; vaccine; tumour specific immunogenic response inducer;
KW astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma, oligodendroglioma,
KW ependyoma, medulloblastoma, primitive neural ectodermal tumour.
XX
OS Homo sapiens
OS Synthetic.
XX
PN CA2290722-A1.
XX
PD 08-JUN-2001.
XX
PF 08-DEC-1999; 99CA-2290722.
XX
PR 08-DEC-1999; 99CA-2290722.
XX
PA (NOVO-) NOVOPHARM RTOFTCH INC.
XX

PI Kaplan HA, Maiti PR, Fast PG, Herman W, Fan MP, Lewis FE;
PI Entwistle JM, MacDonald GC;
XX
XX WPI; 2001 425937/46.
DR
XX
PT Composition useful for treating and diagnosing cancer, comprises stress
PT protein-peptide complexes associated with tumor, and isolated
PT antigen-binding fragments of an antibody that binds specifically to the
PT complex
XX
XX Example 4; Page 108; 154pp; English.
PS
XX
CC The present invention describes a composition (I) comprising stress
CC protein-peptide complexes (SPPC) associated with tumors that is
CC specifically immunogenically cross-reactive with cell surface associated
CC SPPCs specific to target cancer (TC). Also described is an isolated
CC antigen-binding fragment of an antibody that binds specifically to SPPCs
CC or a population of different SPPCs consisting of immunogenic cancer cell
CC surface-associated SPPC of TC. (I) has cytostatic activity and can be
CC used in vaccine production and as a tumour-specific immunogenic response
CC inducer. (I) is useful for treating 71 types of cancers or tumours in a
CC subject, such as astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma,
CC oligodendroglioma, ependymoma, medulloblastoma, and primitive neural
CC ectodermal tumour (PNET). (I) is useful as cancer immunogen including
CC vaccines. (I) is useful for diagnostic and palliative use, for detecting
CC or imaging cancer cells, and to monitor the course of amelioration of
CC malignancy in an individual. AAM43707 to AAM47109 represent peptides
XX which are used in the exemplification of the present invention.
XX
SQ Sequence 7 AA;

Query Match 100.0%, Score 6, DB 22; Length 7;
Best Local Similarity 50.0%; Pred. No. 7.8e+05;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|
Db 1 WRRWNF 6

RESULT 11
AAR57399
ID AAR57399 standard; Protein; 8 AA.
XX
AC AAR57399;
XX
DT 21-MAR-1995 (first entry)
XX
DE Peptide for treating diseases related to anti-DNA antibodies.
XX
KW Carrier, absorbing agent, treatment, anti-DNA antibody, immunogenic
XX
OS Synthetic.
XX
FN JP06102396-A.
XX
PD 12-JUL-1994.
XX
PF 18-JAN-1993; 93JP-0006098.
XX
PR 30-SEP-1992; 92JP-0261821.
XX
PA (KURS) KUPAPAY CO LTD.
XX
DR WPI; 1994-260510/32.
XX
XX A peptide and an adsorbing agent prep. by immobilising it on a
PT carrier useful for treatment of diseases related to anti DNA
PT antibodies and immune complexes
XX
PS Disclosure; Page 12; 14pp; Japanese.
XX
CC The sequences given in AAR57386-413 are peptides which are all covered

stimulating factor (G-CSF a haematopoietic growth factor and
cytokine that stimulates neutrophil proliferation and differentiation),
including evaluation of many factors thought to influence, and be
influenced by, production of white blood cells, in the development of
compounds that bind to G-CSFR, as reagents for detecting G-CSF receptor
or related receptor on living cells, fixed cells, in biological fluid, in
tissue homogenates or in purified natural biological materials, in situ
staining, fluorescence-activated cell sorting (FACS), Western blotting or
enzyme-linked immunoadsorbent assay (ELISA), in receptor purification or
in purifying cells expressing G-CSFR on the cell surface (or inside
permeabilised cells) as a commercial research reagent for various medical
and diagnostic uses or to treat a disease that would benefit from the
ability to of a compound to mimic the effects of G-CSF in vivo
The compounds bind specifically to G-CSFR and allow for studies of
biological activities mediated by the receptor and for the treatment of
diseases, disorders and conditions that would benefit from activating or
inactivating G-CSFR The present sequence is a G-CSFR binding peptide of
the invention.

Sequence 9 AA;

Query Match 100.0%; Score 6; DB 23; Length 9;
Best Local Similarity 50.0%; Pred No 7.9e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWXP 6
|:|:|:
Db 4 WVGWNP 9

RESULT 14

AAR57400
ID AAR57400 standard; Protein; 10 AA.

AC AAR57400;

DT 21-MAR-1995 (first entry)

DE Peptide for treating diseases related to anti-DNA antibodies

KW Carrier: absorbing agent; treatment; anti-DNA antibody; immune complex

OS Synthetic

PN JP06192290 A.

PD 12-JUL-1994.

PF 18-JAN-1993; 93JP-0006098.

PP 30-SEP-1992; 92JP-0261901.

PA (KIPPS) KIPAPAY CO LTD.

WPI; 1994-260510/32.

A peptide and in assisting agent; prep; by immobilising it on a
carrier - useful for treatment of diseases related to anti-DNA

antibodies and immune complexes

Disclosure; Page 12; 14pp; Japanese

The sequences given in AAR57400 413 are peptides which are all covered
by the claimed generic formula.

H-X-(A-B)n-Y-Z

A - Trp, Phe or a peptide fragment consisting of 2 residues;

B = Trp, Phe, Asn or Glu;

X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
consisting of 2 to 10 residues, provided that at least one of
X or Y are present;

Z - OH or NH2; and

n = 2-5.

These peptides may be immobilised on a carrier in the preparation of an

absorbing agent which may be used in the treatment of diseases related
to anti-DNA antibodies and/or immune complex.

Sequence 10 AA;

Query Match 100.0%; Score 6; DB 15; Length 10;
Best Local Similarity 50.0%; Pred No 9.5e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWXP 6
|:|:|:
Db 3 WVFWMF 8

RESULT 15

AAW62149
ID AAW62149 standard; peptide; 11 AA.

AC AAW62149;

DT 16-SEP-1998 (first entry)

DE Haemophilus influenzae tyrosine tRNA synthetase binding peptide 22.

KW Identification; ligand; biological activity; target-binding;

KW drug screening; library; inhibitory ligand.

OS Synthetic.

OS Haemophilus influenzae.

PN WO9819162-A1.

PD 07-MAY-1998.

PF 31-OCT-1997; 97WO-US19638.

PP 31-OCT-1996; 96US-0740671.

FA (NOVA) NOVAGEN PHARM CORP.

FI Fowler CW, Fielding JA, Hyde P, Lydskjaer ST, Ray PR;

WPI; 1998 272389/24.

Identifying ligands which mediate biological activity of a protein
by identifying target-binding ligands and screening a library for
ligands which inhibit target-binding ligand mediated activity

Example 5; Page 100; 143pp; English.

A method has been developed for identifying a ligand which mediates the
biological activity of a target protein (T) by inhibiting the binding
of (T) to a binding partner. The method comprises: (a) screening a first
combinatorial library comprising first member ligands for binding to the
target-binding ligands (TBLs), to identifying one or more TBLs; (b)
screening a second library comprising second member ligands for the
ability to inhibit the binding of one or more of the TBLs to the target
protein, and so obtaining one or more inhibitory ligands; and (c)
determining which of the inhibitory ligands can mediate a biological
activity of the target protein. The present sequence represents a
potential binding peptide for Haemophilus influenzae tyrosine tRNA
synthetase from an example of the present invention. The method can be
used for identifying drugs which can mediate the biological activity of
a target protein. It can be used to identify the biological activity of
a target protein whose biological function is not known and perhaps
cannot be determined directly. The method can also be used to identify
new inhibitory ligands of specific target proteins. The method provides
high throughput screens which are essentially identical for similar and
dissimilar targets, bypassing the need to develop distinct assays for
biochemically diverse targets.

Sequence 11 AA;

Query Match 100.0% Score 67.1419 Length 11;
Best Local Similarity 50.00, Pred. Neg. 2.2e+00;
Matches 3, Conservative 3, Mismatches 0, Indels 0, Gaps 0;

QY 1 WXXWXP 6
1111
DB 5 WPDWGF 10

Search completed: February 20, 2003, 11:26:32
Job time : 30.333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 20, 2003, 10:41:55 / Search time 6.6667 seconds
(without alignments)
37.329 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 38

Sequence: 1 WXXWHF 6

Scoring table: BLOSUM62DX

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	38	100.0	MTCL_CAUCR	Q45971 caulobacter
2	38	100.0	MTSL_RHIME	Q30569 rhizobium m
3	38	100.0	MTSL_RHIME	Q30570 rhizobium m
4	38	100.0	FM6C_SOYBN	P48628 glycine max
5	38	100.0	Y412_APATH	Q04658 arabidopsis
6	38	100.0	FM6C_HPANA	P48627 brassica na
7	38	100.0	FM6C_ARATH	P46312 arabidopsis
8	38	100.0	CATL_PSEPU	Q59714 pseudomonas
9	38	100.0	SLK1_YEAST	Q12460 saccharomyce
10	38	100.0	NOPS_YEAST	Q12460 saccharomyce
11	38	100.0	NOPS_HUMAN	Q29282 homo sapien
12	38	100.0	Y412_APATH	Q04658 arabidopsis
13	38	100.0	Y412_APATH	Q04658 arabidopsis
14	38	100.0	PLSR_CAFEL	Q30569 rhizobium m
15	35	92.1	LYSA_BROME	P37157 drosophila
16	35	92.1	LYSA_BROME	P37157 drosophila
17	35	92.1	LYSA_BROME	P37157 drosophila
18	35	92.1	LYSA_BROME	P37157 drosophila
19	35	92.1	LYSA_BROME	P37157 drosophila
20	35	92.1	LYSA_BROME	P37157 drosophila
21	35	92.1	LYSA_BROME	P37157 drosophila
22	35	92.1	LYSA_BROME	P37157 drosophila
23	35	92.1	LYSA_BROME	P37157 drosophila
24	35	92.1	LYSA_BROME	P37157 drosophila
25	35	92.1	LYSA_BROME	P37157 drosophila
26	35	92.1	LYSA_BROME	P37157 drosophila
27	35	92.1	LYSA_BROME	P37157 drosophila
28	35	92.1	LYSA_BROME	P37157 drosophila
29	35	92.1	LYSA_BROME	P37157 drosophila
30	35	92.1	LYSA_BROME	P37157 drosophila
31	35	92.1	LYSA_BROME	P37157 drosophila
32	35	92.1	LYSA_BROME	P37157 drosophila
33	35	92.1	LYSA_BROME	P37157 drosophila

34	32	84.2	85	1	WG73_BEMLE	Q05288 mycobacteri
35	32	84.2	103	1	ES6D_MYCTU	Q05440 mycobacteri
36	32	84.2	105	1	NIGM_HUMAN	Q05179 homo sapien
37	32	84.2	109	1	NIGM_PVTIN	Q05174 homo sapien
38	32	84.2	111	1	VPX_HYKER	Q74132 human immun
39	32	84.2	111	1	YFBW_SALTI	P81891 salmonella
40	32	84.2	111	1	YFBW_SALTY	Q52328 salmonella
41	32	84.2	112	1	VFX_HV27A	P24110 human immun
42	32	84.2	112	1	VFX_HV2D1	P17760 human immun
43	32	84.2	112	1	VFX_HV2S1	P19045 human immun
44	32	84.2	112	1	VFX_HV2N2	P05915 human immun
45	32	84.2	112	1	VFX_HV2D0	P06939 human immun

ALIGNMENTS

RESULT 1

MTCL_CAUCR

ID MTCL_CAUCR STANDARD: PPT: 358 AA.

AC Q45971, 30-MAY-2000 (Rel. 39, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Modification methylase Ccrm1 (EC 2.1.1.72) (Adenine-specific

DE methyltransferase Ccrm1) (M Ccrm1).

GN CCRM1 OR CCRM OR CC0378.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI_TaxID=155892;

PN [1]

RF SEQUENCE FROM N.A.

RC STRAIN=CB15N / NA1000;

RX MEDLINE=94118303; PubMed=8289276;

PA Zweiger G, Marczynski G, Shapiro L;

PT "A Caulobacter DNA methyltransferase that functions only in the

PT predivisional cell";

RT J. Mol. Biol. 235:472-485(1994).

RN [2]

PP SEQUENCE FROM N.A.

PC STRAIN=ATCC 19089 / CB15;

PX MEDLINE=9173648; PubMed=11259647;

PA Norman W.C., Feldberg J.P., Alley M.P., Ohta N., Madsen T.R.,

PA Eissen J., Heidelberg J.P., Alley M.P., Ohta N., Madsen T.R.,

PA Potocka I., Nelson W.C., Newton A., Stephens C., Bhadke N.D., Ely B.,

PA DeRoy R.T., Dodson P.J., Durkin A.S., Gwin M.B., Haft D.H.,

PA Kolonay J.F., Smit J., Graven M.B., Khouri H., Shetty J., Perry K.,

PA Unterbach T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

PA Salzberg S.L., Venter A.C., Shapiro J., Fraser C.M.;

PT "Complete genome sequence of Caulobacter crescentus";

PL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

CC FUNCTION: THIS METHYLASE REGULATES THE TOPOLE CRAFTED SEQUENCE

CC GANT AND CAUSES SPECIFIC METHYLATION ON A 5' EN WITH STRAINS.

CC APPARAS TO CONTRIBUTE TO THE ACCELERATE CELL-CYCLE CONTROL OF DNA

CC REPLICATION AND CELLULAR MORPHOLOGY.

CC CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine - S-

CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.

CC This Swiss-Prot entry is copyright. It is provided through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL database.

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch>).

CC or send an email to license@sib-sib.ch.

CC EMBL: U01032; AAA18913.1; -

CC EMBL: AE005711; AA02365.1; -

CC HSEF: F14402, 1800.

CC REBASE: 2539; M.Ccrm1.

CC TIGR: CC0378; -

[illegible]

Country	Year	Population	Area	Population Density
Canada	1981	24,000,000	9,970,670 sq. km.	2.4
United States	1981	226,000,000	9,833,517 sq. km.	23.1
France	1981	56,000,000	640,800 sq. km.	87.5
Germany	1981	61,000,000	357,021 sq. km.	171.1
Japan	1981	123,000,000	377,915 sq. km.	325.7
China	1981	1,058,000,000	9,596,961 sq. km.	110.3
India	1981	853,000,000	2,973,147 sq. km.	287.0
U.S.S.R.	1981	246,000,000	17,098,242 sq. km.	14.4
U.K.	1981	56,000,000	244,818 sq. km.	228.7
Italy	1981	56,000,000	301,330 sq. km.	185.8
Spain	1981	40,000,000	505,992 sq. km.	79.1
Sweden	1981	8,500,000	449,964 sq. km.	18.9
Norway	1981	4,500,000	385,207 sq. km.	11.7
Denmark	1981	5,000,000	43,094 sq. km.	116.0
Poland	1981	33,000,000	312,685 sq. km.	105.5
Czechoslovakia	1981	16,000,000	78,866 sq. km.	202.9
Yugoslavia	1981	23,000,000	101,847 sq. km.	226.2
Romania	1981	23,000,000	231,501 sq. km.	99.3
Bulgaria	1981	10,000,000	110,914 sq. km.	90.1
Greece	1981	11,000,000	131,958 sq. km.	83.4
Turkey	1981	56,000,000	783,562 sq. km.	71.5
Iran	1981	40,000,000	1,648,195 sq. km.	24.3
Pakistan	1981	95,000,000	796,095 sq. km.	119.2
India	1981	853,000,000	2,973,147 sq. km.	287.0
China	1981	1,058,000,000	9,596,961 sq. km.	110.3
U.S.S.R.	1981	246,000,000	17,098,242 sq. km.	14.4
U.K.	1981	56,000,000	244,818 sq. km.	228.7
Italy	1981	56,000,000	301,330 sq. km.	185.8
Spain	1981	40,000,000	505,992 sq. km.	79.1
Sweden	1981	8,500,000	449,964 sq. km.	18.9
Norway	1981	4,500,000	385,207 sq. km.	11.7
Denmark	1981	5,000,000	43,094 sq. km.	116.0
Poland	1981	33,000,000	312,685 sq. km.	105.5
Czechoslovakia	1981	16,000,000	78,866 sq. km.	202.9
Yugoslavia	1981	23,000,000	101,847 sq. km.	226.2
Romania	1981	23,000,000	231,501 sq. km.	99.3
Bulgaria	1981	10,000,000	110,914 sq. km.	90.1
Greece	1981	11,000,000	131,958 sq. km.	83.4
Turkey	1981	56,000,000	783,562 sq. km.	71.5
Iran	1981	40,000,000	1,648,195 sq. km.	24.3
Pakistan	1981	95,000,000	796,095 sq. km.	119.2

Y	1	WXXWN:FE	6
		— : : — —	
	2	WXXWN:FE	3
D	3	WXXWN:FE	3

```

RESOUT 2
MTS, RHIVE
ID MTS1, RHIVE STANDARD; PRT; 376 AA.

```

DT 30 MAY-2000 (Ref. 39, Created)
 DT 30-MAY-2000 (Ref. 39, Last sequence update)
 DT 15 JUN-2002 (Ref. 41, Last annotation update)
 DE Modification: methylase Sworf (EC 2.1.1.72) (Adenine-specific
 methyltransferase Sworf) (M_Sworf) (McGinnis).
 DE SWORF OR SWRM OR SWRM2 OR SWRM3
 DT 07 FEBRUARY 2002 (Sinorhizobium meliloti).
 OT Rhizobium meliloti (Sinorhizobium meliloti).
 OT Rhizobium, Rhizobiaceae, alpha-subclass, Rhizobiales,
 Rhizobiaceae, Sinorhizobium.
 OX NCBI_TaxID:362,

20 SCHEEREN, R. H.
 21 STRAIN, J. O.
 22 "Molecular Biology," 1960-1961,
 23 Wright R., Stephens C., Shapiro L.,
 24 "The Cell," 2nd ed., 1960-1961, is published in the third edition
 25 of "Proteobacteria," and its essential components are covered in
 26 "Proteobacteria and Conjugative Proteobacteria",
 27 L. Scheeren, 1960-1961, 1961.

SEQUENCE FROM N.A.
STRAIN-1001,
MEDLINE-0103697, Edited: 1989-07-01, Entry ID: 1001
RA Ochiai D., Bailey, Robert L., Bailey, Robert L., Bailey, Robert L.,
BA Ewald J., Bailey, Robert L., Bailey, Robert L., Bailey, Robert L.,
RA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
BA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
RA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
BA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
RA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
BA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
RA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
BA Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
RT "Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
RT "Goffin, Robert A., Goffin, Robert A., Goffin, Robert A., Goffin, Robert A.,
RL Proc. Natl Acad Sci U S A. 99:9977-9982(2001).

CC - 1. EMOTION. THIS METHOD USES RECOGNITION OF THE STATE RELATED TO THE
CC GAINED AND COGNITIVE RECOGNITION IN A NEW RESEARCH.
CC CORN-MEDIAN METHYLATION HAS IMPORTANT GENETIC RELEVANCE
CC ATTACHED TO THE RECOGNITION OF THE ALTERNATE FORMS OF DNA
CC REPLICATION AND REPAIR MECHANISMS.
CC CATALYTIC ACTIVITY OF ADENOSYL L-GLUTAMINE + GMP ADENOSYL
CC ADENOSYL L-HOMOCYSTEINE + GMP + L-GLUTAMINE FOR THE

This SMS-PRO entry is copyright © 2012 and shared through a collaboration between the Swiss Institute of Bioinformatics and the EMBL collaboration. The European Bioinformatics Institute and the EMBL collaboration have not reviewed this submission and do not warrant its use in any way. The EMBL and EBI logos are not allowed to be used by and for commercial entities. Logos are licensed under <http://www.ebi.ac.uk/embldata/logo-usage> or send an email to licensed@ebi.ac.uk

```

DR EMPL AP011804; AAR71360.1; -
DE EXPR_A0501795; SAC45498.1; -
DE ERASE1_1024; M.McElrP.
DE InterPro; IPR011091; CNA_Mettransf.
DE TRNTRF; IPR022225; RNAB6_mtfase.
DR InterPro; IPR002941; N6/N4_Mtasec.
DE InterPro; IPR000051; SAM_bind.
DR InterPro; IPR000051; SAM_bind.
DE Pfam; PF01535; N6_N4_Mtasec; 1.
DE PRINTS; PR001704; ZENITMTAFASE.
DE PRINTS; PF001064; ZENITMTAFASE.
DE PROSITE; PS00092; No_MTASEC; 1.
DM Translasec; MethylTranslasec; DNA_replication; Complete_protome
EM CONFLICT 135 141 REVERSE_SEQUENCE (IN REF. 1).
FT CONFIDAT 157 159 FROM A (IN PPT 1).
DE TRANS 100 AA; 4041; 90; 1155511100A CRCE4;

```

Query Match	100.0%	Score 38;	DB 1;	Length 376;
Best Local Similarity	66.7%	Prod. No. 52;		
Matched	4;	Conservative	0;	Models 0;
			Wgs/seq/line	0;

QY	1	WXXWHF	6
		:	
Db	347	WTFWHF	352

RESULT 3	
MTB1_BRUB	
ID MTB1_BRUB	STANDARD; PRT; 377 AA

OT 30 MAY 2006 (Ref. 39, Created)
 OT 31 MAR 2006 (Ref. 39, Last database update)
 OT 16 OCT 2001 (Ref. 40, Last annotation update)
 TE Multitaxa (multitaxa Edit: 1992.01.02) (Adaptive specific
 OR methyltransferase BabI) (M.BabI) (M.Cormi).
 GN BabIM OR CCRM.
 OS Brucella abortus.
 OC Brucella, Tetracyclorhiza, alpha-Proteobacteria, Rhizobiaceae group.
 OC Brucellaceae, Brucella.
 OC alpha-Proteobacteria;
 OC Rhizobiaceae;

RE REFERENCE FROM HIM.
 EC STEIN=82308;
 EX METLINE-973440129; PubMed-0294447;
 FA "Mihailo E. Stephens C. Shapiro T."
 KI "the Csm Gm Activity Inactivation is widespread in the alpha subgenus of
 RT of Proteobacteria, and its essential functions are conserved in
 RI Enkadium m-11041 and Gm-14471 of Proteobacteria";
 RI J. Eukaryot. Microbiol. 1997;44(3):197.

1. FUNCTION. THIS METHYLASE RECOGNIZES THE FOLLOWING SEQUENCES
 2. 5'-GATC-3' AND CACCG-3' AND CATALYZES METHYLATION ON A C IN BOTH STRANDS.
 3. 4. DNA METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
 4. 5. IT APPEARS TO CONTRIBUTE TO THE ACCURATE CELL-CYCLE CONTROL OF DNA
 5. 6. REPLICATION AND CELLULAR MORPHOLOGY.
 6. 7. CATALYZES ACTIVITY OF 5-Methylthioadenosine + DNA adenine = S-
 7. 8. adenosyl-L-homocysteine + DNA C-methylaminopurine.

On 21 May 1967, the Soviet Union announced that it had signed a "collaboration agreement" with the European Economic Community and the EEC's laboratory for the European Radiobiological Institute. There are no restrictions on its use by non-EEC countries, and it is open to all countries in the "commercial" category and all of the states of the "free world" and "non-aligned" countries. The agreement is a 10-year agreement. The EEC will provide equipment and staff and provide a laboratory (see page 17) which will be announced or send an appeal to the Soviet Union.

CC	EMBL; AF011805; AF071251.1; -
DE	PPRASE; 30kD; M Bahl
PP	EMBL; AF001031; GenBank; 26688
PS	EMBL; AF001031; GenBank; 26688
TE	EMBL; AF001031; GenBank; 26688
TR	EMBL; AF001031; GenBank; 26688
TS	EMBL; AF001031; GenBank; 26688
TT	EMBL; AF001031; GenBank; 26688
DR	EMBL; AF001031; GenBank; 26688

DR PRINTS: PR00506; D21N6MTPPASE.
DR PRINTS: PR00508; S21N4MTPPASE.
DR PROSITE: PS00092; N6_MTPASE; 1.
KW Transferase; Methyltransferase; DNA replication.
SQ SEQUENCE 377 AA; 40000 MW; 6570RAA256R0R3P CFC64.

Query Match 100.0%; Score 39; DB 1; Length 377;

Best Local Similarity 66.7%; Pred No 52;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWHF 6
Db 347 WTFWHF 352

RESULT 4

FDEC SOYBN

ID_FDEC SOYBN STANDARD, FET, 424 AA
AC P48628;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Omega-6 fatty acid desaturase, chloroplast precursor (EC 1.14.99.-).

OS Glycine max (soybean)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;

OC Eurosids II; Fabiales; Fabaceae; Papilionoideae; Phaseolaceae; Glycine

OX NCBI_TaxID=3847;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Seed;

RX MEDLINE=4345609; PubMed 464123;

RA Hiltz W.D., Carlson T.J., Booth J.P., Jr., Kinney A.J., Stecca P.L.,

RA Yadav N.S.;

RT "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA

and its expression in a cyanobacterium";

RL Plant Physiol. 105:635-641(1994).

CC - FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES

THE SECOND DOUBLEBOND IN THE BIOSYNTHESIS OF 18:3 AND 18:4 FATTY

ACIDS. IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT

TO USE FERRICYTOCHROME AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS

ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLALCOHOL

CC - PATHWAY: Polyunsaturated fatty acid biosynthesis.

CC - SUBCELLULAR LOCATION: CHLOROPLAST; MEMBRANE POUND (PROBABLY).

CC - DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE

AND/ OR BE INVOLVED IN METAL ION BINDING

CC - SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch)

CC -----

CC EMBL: L29215; AAA0156 1;

DR InterPro: IPR01225; FA desaturase.

DR Pfam: PF00487; FA desaturase; 1.

DR ProDom: PD001081; FA desaturase; 1.

KW Oxidoreductase, Fatty acid biosynthesis; Chloroplast; Membrane;

KW Transit peptide.

FT TRANSIT 1 63 CHLOROPLAST (BY SIMILARITY).

FT CHAIN 64 424 OMEGA 6 FATTY ACID DESATURASE

FT DOMAIN 165 169 HISTIDINE BOX-1.

FT DOMAIN 201 205 HISTIDINE BOX-2

FT DOMAIN 361 365 HISTIDINE BOX-3.

SQ SEQUENCE 424 AA; 49641 MW; 4FHFQDAP1A2009 CFC64.

Query Match 100.0%; Score 39; DB 1; Length 424;

Best Local Similarity 66.7%; Pred No 59;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0.

OY 1 WXXWHF 6
Db 251 WLMWHF 256

RESULT 5

Y412 ARATH

TC_Y412 ARATH STANDARD, FET, 439 AA.
AC C04658;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein At5g27120.

GN AT5G27120 OF TM021804 12 OF T2184_10.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;

OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016721; PubMed-11139714;

RA Tabata S., Kaneke T., Nakamura Y., Katani H., Kato T., Asamizu E.,

RA Miyajima H., Sasamoto M., Kimura T., Haseguchi T., Kawashima K.,

RA Kohara M., Katsumoto M., Matsuno A., Muraki A., Nakayama S.,

RA Nakazaki N., Naruo K., Okumura K., Shimizu S., Takayuchi S., Wada T.,

RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,

RA Huang B., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,

RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,

RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,

RA Belter E., Cordum H., Cordus W., Courtney L., Courtney W., Dante M.,

PA Du H., Edwards J., Fryman J., Hakonen R., Lamy E., Lattrell P.,

PA Legendre S., Meyer P., Mulvaney E., Owersky P., Riley A., Stromatt C.,

PA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,

PA Parnell L., Shah P., Rodriguez M., Hoon See L., Vil D., Baker D.,

RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Maira M.,

PA Martensen F., McCobbie W.P., Wilson P.K., Murphy G., Bancroft I.,

PA Volokhov G., Winkler F., Dostert-Hoff A., Sticker W., Fohl T.,

PA Erdman K.D., Tealyn N., Hartley N., Bent E., Johnson S.,

PA Langham S.A., McWilliams B., Folken S., Grygier P., Zimmermann W.,

PA Ramsberger U., Wedler H., Balke K., Wedler E., Peters S.,

PA Wei-Zeigler T., Roche G., Fise M., Hanf J., Reinhardt G., Hoyer S.,

PA Feldausch M., Lamrecht S., Villaverde E., Stelzer G., Atiles W.,

PA Ben S.G., Lerche K., Klesner G., Mayer K., Frits S., Sch. of H.,

PA Schueller C., Zaccaria P., Mawes H.-W., Revan M., Franz P.,

RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis

thaliana";

RL Nature 409:823-926(2000).

CC - SIMILARITY: BELONGS TO THE NDE/NOP6 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch)

CC -----

CC EMBL: AF067271; AAB61073.1; -

DR InterPro: IPR02687; Nop

DR Pfam: PF01798; Nop; 1.

DR ProDom: PD004104; Nop; 1

KW Hypothetical protein.

SQ SEQUENCE 439 AA; 47932 MW; BE0ED214F9EC4FDB CFC64;

Query Match 100.0%; Score 39; DB 1; Length 439;

Best Local Similarity 66.7%; Pred No 59;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWHF 6

Db 251 WLMWHF 256

Query Match 100.0%; Score 39; DB 1; Length 439;

Best Local Similarity 66.7%; Pred No 59;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWHF 6

Db 251 WLMWHF 256

Query Match 100.0%; Score 39; DB 1; Length 439;

Best Local Similarity 66.7%; Pred No 59;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

[illegible][illegible]

CC THE SECOND FOURTEENTH IN THE BIOSYNTHESIS OF 14:3 AND 14:3 FATTY
CC ACIDS, IMPORTANT COMPONENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC TO BE DIFFERENTIAL AS AN ELECTRON DUMP AND TO ACT ON FATTY ACIDS
CC ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL
CC PATHWAY. POLYUNSATURATED FATTY ACID BIOSYNTHESIS
CC STRUCTURAL LOCATION: CHLOROPLAST, MEMBRANE-BOUND (APPARABLE)
CC -1- DEVELOPMENTAL STAGE: HIGHEST LEVELS FOUND IN EXPANDING LEAVES.
CC -1- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/ OR BE INVOLVED IN METAL ION BINDING.
CC -1- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>
CC or send an email to license@isb.sib.ch)

DR EMBL; U09503; AAA92800.1; -
DR EMBL; AL022198; CAA18198.1; -
DR EMBL; AL161578; CAB79813.1; -
DR EMBL; AY045621; AAK73979.1; -
DR EMBL; AY058078; AAL24186.1; -
DR EMBL; AY059852; AAL24240.1; -
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR ProDom; PD001081; FA_desaturase; 1.
KW Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW Transit peptide.
FT TRANSIT 1 69 CHLOROPLAST (BY SIMILARITY)
FT CHAIN 70 448 OMEGA-6 FATTY ACID DESATURASE.
FT DOMAIN 171 175 HISTIDINE BOX-1.
FT DOMAIN 207 211 HISTIDINE BOX-2.
FT DOMAIN 367 371 HISTIDINE BOX-3.
SQ SEQUENCE 448 AA; 61395 MW; C2AC72F828F8297 CFC64;

Query Match 100.0%; Score 38; DB 1; Length 448;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWHF 6
DB 297 WXXWHF 262

RESULT 8
CATA_PSEPU STANDARD; PRT; 479 AA.
AC Q59714;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA OP CATA.
OS Pseudomonas putida
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadiales;
OC Pseudomonas;
OX NCBI_TaxID=303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Corvallis;
RX MEDLINE=98019091; PubMed-9358059;
RA Kim Y C, Miller C D, Anderson A J.;
RT "Identification of adjacent genes encoding the major catalase and a
RT bacterioferritin from the plant-beneficial bacterium Pseudomonas
RT putida";
RL Gene 1991219-224(1997)
CC
CC FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND, WHEN, SEVERE
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE
CC -1- CATALYTIC ACTIVITY: 2 H2O2(O2) -> O(2) + 2 H2O(O).
CC -1- COFACTOR: HEME GROUP.
CC -1- ENZYME REGULATION: ACTIVATED BY PEROXIDE.

CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>
CC or send an email to license@isb.sib.ch)

DR EMBL; U63511; AAB88219.1; -
DR HSSP; F42321, CCAB.
DR InterPro; IPR002226; Catalase.
DR Pfam; PF00193; Catalase; 1.
DR EPRINTS; PR00067; CATALASE.
DR ProDom; PD000510; Catalase; 1.
DR PROSITE; PS00437; CATALASE_1; 1.
DR PROSITE; PS00438; CATALASE_2; 1.
KW Oxidoreductase; Peroxidase; Iron, Heme, Hydrogen peroxide.
FT ACT_SITE 53 53 BY SIMILARITY.
FT ACT_SITE 126 126 BY SIMILARITY.
FT BINDING 336 336 PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 479 AA; 53381 MW; EEE3CBDE67778571 CFC64;

Query Match 100.0%; Score 38; DB 1; Length 479;
Best Local Similarity 66.7%; Pred. No. 64;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWHF 6
DB 209 WXXWHF 214

RESULT 9
SIKI_YEAST STANDARD; PRT; 504 AA.
ID SIKI_YEAST
AC Q12460;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SIKI protein (Nucleolar protein NOP56).
GN SIKI OR NOP56 OR YLR197W OR L8167.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / YPH1;
RX MEDLINE=96040178; PubMed=7547500;
RA Morin F.U., Downs J.A., Snodgrass A.M., Gilmore T.D.;
RT "Genetic analysis of growth inhibition by GATA-1 kappa P-alpha in
RT Saccharomyces cerevisiae";
FI Cell Growth Differ. 6:789-798(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Rinkman F., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Grew T., Kirsten J., Kuzhna T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille F., Mardis E., Meneses S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rinken L., Riles L.,
RA Tach S., Trevaekis F., Vaudin M., Vignati P., Wilcox E., Wilson R.,
RA Wohlman P., Waterston R.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBP databases.
CC
CC [3]
CC CHARACTERIZATION, AND MUTAGENESIS.
RP MEDLINE=98038777; PubMed-9372940;
FA Jallier T., Bujes T., Tollevey C., Hurt E.;
RT "Nucleolar PHE/D repeat proteins Hsp56 and Hsp59 interact with Nop1p
RT and are required for ribosome biogenesis";
RL Mol Cell Biol 17:7088-7098(1997).
CC
CC FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.

```

CC -1- SUBMITTING INSTITUTE WITH RELEASE RIGHTS:
CC THE SWISS INSTITUTE OF BIOINFORMATICS
CC -1- SIMILARITY BELONGS TO THE NIPS/NIPFS FAMILY
CC -----
CC THIS SWISS PROT ENTRY IS STEADYFAST. IT IS PROVIDED THROUGH A NON-REPRODUCTION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OPERATION
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY ANY PROFIT INSTITUTIONS AS LONG AS THE INTENT IS NOT AT ALL
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sdb.org/
CC OR SEND AN EMAIL TO license@isb-sdb.ch).
CC -----
CC DR EMBL_012637; AAC948366.1; -
CC DR EMBL_014413; AAB61491.1; -
CC DR SGI_2001157; SIRT.
CC DR InterPro IPR002697; NOP.
CC DR ELM_012115; NOP.
CC DR ELM0001; ELM04104; NOP.
CC KM Ribosome Biogenesis, nuclear protein.
CC FT DOMAIN 443 504 ASP/GLU/LYS-RICH.
CC FT MOLISEN 333 333 V-A: REDUCED GROWTH RATE AT ALL
FT TEMPERATURES, WHEN ASSOCIATED WITH E-3B5.
FT MOLTAGEN 355 355 T-45: AT 37 DEGREES, GROWTH SLOWS AFTER 6
FT MOLTAGEN 355 355 TO 6 HOURS AND CELL DIVISION STOPS AFTER
FT 20 HOURS.
FT MOLTAGEN 365 365 M-SR: REDUCED GROWTH RATE AT ALL
FT SEQUENCE 501 AA, 50804 NM, F022A5970B19493 CNOG4,
SQ TEMPERATURES, WHEN ASSOCIATED WITH A 333.
Query Match Score 38, DB 1, Length 504
Best Local Similarity 66.1%, Ref. No. 57,
No-desc 1, Coverage 100%, Weight-Bias 1, Weight 1, Type 1,
DY 1 WXXWHF 6
Db 199 WXXWHF 203
RESULT 10
NOPFS YEAST STANDARD PRT 511 AA.
AC 112459:
DT 01-NOV-1997 (Ref. 35, Created)
DT 01-NOV-1997 (Ref. 35, Last sequence update)
DT 15-JUN-2002 (Ref. 41, Last annotation update)
DE Nucleolar Protein NOPFS (M. 112459; pI 4.14; pI 4.14; pI 4.14).
GN Nucleolar Protein of Yeast (NOPFS).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi, Ascomycota, Basidiomycota, Saccharomycetes.
OC Saccharomycetidae, Saccharomycetales, Saccharomycetaceae.
OX NC_012459.1.
RN 1.
RP SEQUENCE FROM N.A.
RC STRAIN S288c / FY1679,
RA Pearson R.M., Hernandez Y., Mills J.S., Hingley-Parker A., and others (1997),
RI "The Yeast Genome Database: The FYPO/Jenyon/Tiroc Database."
RF 121
RE SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC STRAINW333.
RX MEDLINE 3632916, ELM01 3632916.
RA W.F. Piotti, J.H. Lee, W. Hoffmann, A. Chou, S. Alt, J.
RT "NOPFS is a small nucleolar ribosomal protein component involved in
RI pre-18S rRNA processing in yeast."
RI C. Biol. Chem. 273:16453-16458(1998).
CC -1- FUNCTION: REQUIRED FOR PRE-18S RNA PROCESSING. MAY BIND
CC MICRORNAs
CC -1- SUBJECT: INTERACTS WITH RPLP AND RPLP.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE NIPS/NIPFS FAMILY
CC -----
CC THIS SWISS PROT ENTRY IS STEADYFAST. IT IS PROVIDED THROUGH A NON-REPRODUCTION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OPERATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS

```

CC BY-NC-ND 4.0 International license.

The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

or send an email to license@lsb.sib.ch.

NP FWERL X986F5, CAA02165.1; -
DE EMBL Z75217, CAA96830.1; -
DE EMBL AF020170, AAC09484.1; -
DE SDB S0055937, NOP58.
LA InterPro, IPRO02587, NBP.
DR RefSeq, EF01708, Nbp_1.
PM ProDom, PD004104, Nop; 1.
KW Ribosome biogenesis, Nuclear protein, RNA processing,
LM LAMAIN 441
BI Sequence 511 AA, 5456 MW, 8A06A444P, NTEN CF764;
SI

Query Match 100%, Score 59, EE 1, Length 511;
Best local similarity 66%, Field No. 68;
Matrix 4, Consistency 0, Mismatch 0, Indels 0, Gaps 0

QY I WXXWVF 6
|||
-- 167 WXXWF 102

RESULT 11
NOP5 HUMAN STANDARD PPT 529 AA.

AC Q9VZX3, Q9FOG6, Q9OFN3;
DT 15 OCT 2001 (Rel. 40, Created)
DI 19-03-2001 (Rel. 40, Last sequence update)
FT FTCD 2001 Feb 4; last annotation update)
TT Nucleolar protein NOP5 (nucleolar protein 5) (NOP58) (HSPC120).
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniota, Vertebrata, Euteleostomi;
OC Mamalia, Eutheria, Plimates; Catarrhini, Hominoidea.
OX NCRI_TaxID=9606;

RN [1]
RF SEQUENCE FROM N.A.
RA Lyman S.K., Gorace L.;
RT Cloning and characterization of NOP5/NOP58";
FI Submitted (JAN-1999) to the FWERL/GenBank/EBI databases.
RN [2]
RF SEQUENCE FROM N.A.
RA MERITINO J994673, PUBMED 10995055;
FA Polson G.A., Smith D.P.F., Tirochelle W.D.;
ET Identification and characterization of a novel progf-induced human gene.";
FI Genes Dev 11: 1199-1209.

RN [3]
RF SEQUENCE OF 1 447 FROM N.A.
RC TISSUE=Brain;
RA Bloecher H., Boecker M., Brandt F., Mewes H.-W., Gassenhuber J.,
RA Wiemann S.;
LI Journal Biol Chem 274(19): 12993-12998 FEB 1999; PubMed/PMC databases.
RN [4]
RF SEQUENCE OF 54 523 FROM N.A.
RC TISSUE=Blood;
HA MH0109449967, PubMed 11642123;
EA Chang C.H., Ye K.Y., Ma X.T., Fan S.X., Zhu M., Chen S.J., Fu S.Y.,
FA Guan Y., Fu B.T., Li C., Zheng Y., Xu X.S., Han Z.G., Zhang C.W.,
FA Tang W., Huang C.H., Zhou G., Hu G.X., Gu J., Chen S.J., Chen Z.,
ET Cloning and functional analysis of CDNA with open reading frames for
BT 353 previously unidentified genes expressed in CD34+ hematopoietic
AT Stem/progenitor cells.";
RL Genome Res 10:1546-1560(2000).
CL FUNCTION RESEMBLES THE 52S RIBOSOMAL PROTEIN EUCOELESTIC BY
CM SIMILARITY).

CC CELLULAR LOCATION Nuclear, nucleolar.
CC TISSUE SPECIFICITY Ubiquitous.
CC ORGANISM HOMO SAPIENS IN THE NOP5/NOP58 FAMILY.
CC THIS Swiss-Prot entry is copyrighted and produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: AF123534; AAD27610.1; -
DR EMBL: AF263608; AAF61344.1; -
DR EMBL: AL117554; CAB55989.1; -
DR EMBL: AF161469; AAF29094.1; -
DR SWISS-2DPAGE; Q9YEX3; HUMAN.
DR InterPro; IPR002687; NTP.
DR Pfam; PF01798; Nop; 1.
DR ProDom; PD004104; Nop; 1.
KW Ribosome biogenesis; Nuclear protein.
FT CONFLICT 2 2 L -> M (IN REF. 3).
FT CONFLICT 129 129 G -> V (IN REF. 3).
FT CONFLICT 221 221 LTYCYCTQVVDREPRITASAV -> VHTASVPRPLAISPLCT
FT CONFLICT 235 260 CQ (IN REF. 4).
FT CONFLICT 280 280 KAAAEISMGTEVEEDICNIIILCTQ -> EGSCPDHGNP
FT CONFLICT 443 447 M -> V (IN REF. 3).
FT CONFLICT 447 447 PTIEQ -> TQNTT (IN REF. 3).
SQ SEQUENCE 529 AA; 54579 MW; 2702730PF69A5EC CP064;

Query Match 100.0%; Score 38; DB 1; Length 529;
Best Local Similarity 66.7%; Pred. No. 70;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 186 WYGMHF 191

RESULT 12
NOP5_RAT STANDARD; PRT; 534 AA.
AC Q9QZ86; O88525;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleolar protein NOP5 (Nucleolar protein 5) (Nop140 associated
DE protein).
GN NOP5 OP NAB65.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX NCBI_TaxID 10116;
RN [1]
RF SEQUENCE FROM N.A. : SEQUENCE OF 1 53, AND INTERACTION WITH NOP1
RX MEDLINE 2014370, PubMed 10672015;
RA Yang Y, Isaac C, Wang C, Dragon F, Pogacic V, Meier T;
RT "Conserved composition of mammalian box H/ACA and box C/D small
RT nucleolar ribonucleoprotein particles and their interaction with the
RT common factor Nop140.";
RL Mol. Biol. Cell 11:567-577(2000).
RN [2]
RP SEQUENCE OF 1-461 FROM N.A.
RA Hutton D, Gray J.C.;
RU Submitted (JUN 1998) to the EMBL/GenBank/DBPJ databases.
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT BIOGENESIS (BY
CC similarity).
CC -1- SUBUNIT: Interacts with Nop1/Nop140.
CC -1- SUBCELLULAR LOCATION: Nuclear, nucleolar.
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF194371; AAF05769.1; -
DR EMBL: AF069782; AAC23535.1; -
DR InterPro; IPR002687; Nop.
DR Pfam; PF01798; Nop; 1.
DR ProDom; PD004104; Nop; 1.
KW Ribosome biogenesis; Nuclear protein.
FT CONFLICT 306 306 E -> K (IN REF. 2).
FT CONFLICT 459 459 A -> K (IN REF. 2).
SQ SEQUENCE 534 AA; 60070 MW; 4B9585FA14E67739 CRC64;

Query Match 100.0%; Score 38; DB 1; Length 534;
Best Local Similarity 66.7%; Pred. No. 71;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 521 WDFWHF 526

RESULT 13
Y4CD_RHISN STANDARD; PRT; 640 AA.
ID Y4CD_RHISN
AC P55386;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 71.6 kDa protein Y4CD.
GN Y4CD.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym PNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phizobiaceae; Phizobium
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:97305956; PubMed=9163424;
RA Freilberg C A, Fellay P, Raitoch A, Proughon W J, Rosenthal A,
RA Perret X,
PT "Molecular basis of symbiosis between Phizobium and legumes."
RL Nature 387 394-401(1997).
RN [1]
CC -1 SIMILARITY: NONE OBVIOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: AEO0067; AAB91634.1; -
DR InterPro; IPR00157; TIF_domain.
DR SMART; SMO0255; TIF; 1.
KW Hypothetical protein, Plasmid.
SQ SEQUENCE 640 AA; 71604 MW; 814F000000A000F CF064;

Query Match 100.0%; Score 38; DB 1; Length 640;
Best Local Similarity 66.7%; Pred. No. 84;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 521 WDFWHF 526

RESULT 14
PLSB_CAEEL STANDARD; PRT; 718 AA.
ID PLSB_CAEEL
AC Q22949;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DR PIR: S32726; S32726.
 DR HSSP; P00695; 1LZ5.
 DR FLYBase; FBgn0011201; LysA.
 DR FLYBase; FBgn0004426; LysC.
 DR FLYBase; FBgn0004427; LysD.
 DR InterPro; IPR001916; GH_22.
 DR Pfam; PF00062; Lys; 1.
 DR PRINTS; PR00135; LYZLACT.
 DR SMART; SM00263; LYZ1; 1.
 DR PROSITE; PCS0128; LACTALBUMIN_LYSOZYME; 1.
 KW Hydroxylase; Glycosidase; Bacteriolytic enzyme; Signal;
 KW Multigene family.
 FT SIGNAL 1 18 BY SIMILARITY.
 FT CHAIN 19 140 LYSOZYME A/C/D.
 FT DISULFID 24 139 BY SIMILARITY.
 FT DISULFID 45 129 BY SIMILARITY.
 FT DISULFID 80 96 BY SIMILARITY.
 FT DISULFID 92 110 BY SIMILARITY.
 FT ACT_SITE 50 50 BY SIMILARITY.
 FT ACT_SITE 68 68 BY SIMILARITY.
 FT CONFLICT 41 42 AP -> NK (IN REF. 1);
 SQ SEQUENCE 140 AA; 15635 MW; 75C24CA6F85DF903 CRC64;

Query Match 92.18; Score 35; DB 1; Length 140;
 Best Local Similarity 50.08; Pred. No. 56;
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXWHF 6
 DB 123 WSTWHY 128

Search completed: February 20, 2003, 10:48:03
 Job time : 7.66667 secs



Genome version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein Protein search using SW model

Run on February 20, 2003, 10:45:10, Search time: 0.003 seconds
(without alignment)
52.984 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 38

Sequence: 1 WXXWHF 6

Gapopen: 10.0, Gapext: 0.5

Searched: 671580 seqs, 286347115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL_21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriophage:
17: sp_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	167	Q9CLS4	Q9CLS4 Pasteurella
2	38	100.0	160	Q96N05	Q96N05 Homo sapien
3	38	100.0	201	Q9C600	Q9C600 Arabidopsis
4	38	100.0	202	Q9AXY3	Q9AXY3 Brassica ca
5	38	100.0	213	Q9AXY5	Q9AXY5 Brassica na
6	38	100.0	213	Q9AXY3	Q9AXY3 Brassica na
7	38	100.0	219	Q94K47	Q94K47 Arabidopsis
8	38	100.0	232	Q9AXY4	Q9AXY4 Brassica na
9	38	100.0	230	Q94K47	Q94K47 Arabidopsis
10	38	100.0	241	Q99123	Q99123 Mycobacteri
11	38	100.0	256	Q9FIN2	Q9FIN2 Arabidopsis
12	38	100.0	260	Q93250	Q93250 Treponema p
13	38	100.0	289	Q8TBL3	Q8TBL3 Methanocarc
14	38	100.0	293	Q91065	Q91065 Pseudomonas
15	38	100.0	310	Q9EXF1	Q9EXF1 Listeria mo
16	38	100.0	314	Q91120	Q91120 Streptomyces

17	38	100.0	381	2	Q93TQ4	Q93TQ4 agrobacteri
18	38	100.0	386	16	Q80H89	Q80H89 agrobacteri
19	38	100.0	394	10	Q9AWM9	Q9AWM9 guillardi
20	38	100.0	403	16	Q81FS6	Q81FS6 bruceella se
21	38	100.0	415	16	Q81FS6	Q81FS6 thermocact
22	38	100.0	421	16	Q93E64	Q93E64 E. coli
23	38	100.0	423	16	Q927H4	Q927H4 listeria in
24	38	100.0	423	16	Q97412	Q97412 listeria in
25	38	100.0	424	16	Q94853	Q94853 chlamydomon
26	38	100.0	452	16	Q8ZL19	Q8ZL19 salmonella
27	38	100.0	452	16	Q8ZL19	Q8ZL19 salmonella
28	38	100.0	454	10	Q928W0	Q928W0 cinder arlet
29	38	100.0	462	16	Q9XFT1	Q9XFT1 coccidiophila
30	38	100.0	472	16	Q8VYH8	Q8VYH8 anabaena sp
31	38	100.0	473	5	Q9RH17	Q9RH17 leishmania
32	38	100.0	473	11	Q93426	Q93426 E. coli
33	38	100.0	475	17	Q9HSP8	Q9HSP8 halobacteri
34	38	100.0	485	10	Q80401	Q80401 oryza sativ
35	38	100.0	487	5	Q45012	Q45012 nemorhabdi
36	38	100.0	488	16	Q94SV0	Q94SV0 caulobacter
37	38	100.0	492	16	Q9Y731	Q9Y731 anabaena sp
38	38	100.0	494	17	Q97475	Q97475 sulfolobus
39	38	100.0	497	3	Q94514	Q94514 schizosacch
40	38	100.0	499	10	Q91TV0	Q91TV0 arabidopsis
41	38	100.0	508	3	Q9P757	Q9P757 schizosacch
42	38	100.0	508	17	Q97WR4	Q97WR4 sulfolobus
43	38	100.0	510	5	Q9XW4	Q9XW4 thiosphila
44	38	100.0	510	16	Q9XW4	Q9XW4 thiosphila
45	38	100.0	511	5	Q9VMA6	Q9VMA6 thiosphila

ALIGNMENTS

RESULT 1

Q9CLS4

ID Q9CLS4 PRELIMINARY; PRT; 187 AA.

AC Q9CLS4;

DT 01-JUN-2001 (Trembl, 17, Created)

DT 01-JUN-2001 (Trembl, 17, Last sequence update)

DT 01-MAR-2002 (Trembl, 20, Last annotation update)

DE Hypothetical protein PM1135.

GN PM1135.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

CC Pasteurella.

OX NCBI_TaxID=747;

RN [1]

RE SEQUENCE FROM N.A.

FC STRAIN=EM70;

PX MEDLINE:114586, PubMed=11248100;

RA May R.J., Zhang Q., Li L.B., Paustian M.L., Whitam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida pm70.";

PL Proc Natl Acad Sci U S A 98:3460-3465(2001).

CC -1- SIMILARITY: BELONGS TO THE TETR/ACF FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

DR EMBL: AB006154; MAF02219.1;

DR InterPro: IPR01647; HTH_Tetr.

DP Pfam: PF00440; Tetr_1.

DR PRINTS: PR00455; HTHETR.

KW DNA-binding, Hypothetical protein, Transcription regulation;

KW Complete proteome.

SQ SEQUENCE 187 AA; 2225 MW; 482399534111F (RC064);

Query Match: 100.0%, Score 38, SP 16, Length 187;

Best Local Similarity: 66.7%; Pred. No. 1.9e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:|:|
Db 89 WNTWHF 94

Db 154 WXXWHF 162

RESULT 6

ID Q9AXY2 PRELIMINARY, FFT, 216 AA.

AC Q9AXY2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Omega 6 reductase (Fragment).

GN FAD6-BN-1.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QV, STELLAR;
RA Fourmann M., Froger N., Brunel D.;
RT "Amplified consensus gene markers: Tools designing for a genetic map
of Arabidopsis-known-function genes in Brassica";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases
DR EMBL; AF229391; AA00663.1;
DR InterPro; IPR001225; FA_desaturase
DR Pfam; PF00487; FA_desaturase; 1.
DR ProDom; PD001081; FA_desaturase; 2.
FT NON_TER 1 1
FT NON_TER 216 216
SQ SEQUENCE 216 AA; 25110 MW; EFFQEDDEBIBDEB3E4 CRO64;

Query Match 100.0%; Score 39; DB 10; Length 216;
Best Local Similarity 66.7%; Pred No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
Db 154 WXXWHF 162

RESULT 7
ID Q94K47 PRELIMINARY, FFT, 219 AA.
AC Q94K47;

DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative 6-phosphogluconolactonase
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.Y., Sakano H., Pham P.V., Park J., Chung M.Y.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C.C., Toriumi M., Yin G.,
RA Bowers J., Carninci P., Chen H., Cheu P., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai T., Kim C., Kossema F.,
RA Lam P., Lin J., Meyers R.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Sato M., Seki M., Shim P., Southwick A.,
RA Shinozaki K., Davis P.W., Ecker J.R., Theologis A.,
RT "Full Length cDNA of gene F16H17.11/ATG24400 (G11017901)";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases
CC - SIMILARITY CONTAINING 1 WS EMBL (EFF A67 COMAIN).
DP EMBL; AF370305; AA044120.1;
DR TIGRFAMs; TIGR01198; pgl; 1.
KM Repeat; WD Repeat;
SQ SEQUENCE 219 AA; 24542 MW; 1904A723D24348RA CRO64;

Query Match 100.0%; Score 39; DB 10; Length 219;

Best Local Similarity 66.7%; Pred No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
Db 72 WARWHF 77

RESULT 8

ID Q9AXY4 PRELIMINARY, FFT, 222 AA.
AC Q9AXY4;

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Omega 6 reductase (Fragment).

GN FAD6-BN-1.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QV, RAPID CYCLING;
RA Fourmann M., Froger N., Brunel D.;
RT "Amplified consensus gene markers: Tools designing for a genetic map
of Arabidopsis-known-function genes in Brassica";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229391; AA00661.1;
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR ProDom; PD001081; FA_desaturase; 2.
FT NON_TER 1 1
FT NON_TER 222 222
SQ SEQUENCE 222 AA; 25011 MW; ADF14F44444444 CRO64;

Query Match 100.0%; Score 38; DB 10; Length 222;
Best Local Similarity 66.7%; Pred No. 2.2e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
Db 163 WXXWHF 168

RESULT 9

ID Q53604 PRELIMINARY, FFT, 230 AA.
AC Q53604;

DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein Rv0059.
GN Rv0059 CP MT0002 02 CP MT0005.

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Smith T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles C., Harpin R., Holtz E.,
RA Hornsby T., Jagels K., Krogh A., McLean L., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
RA Ruston J.E., Taylor K., Whitehead S., Barrall H.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";

DB 220 WEPWHF 225

RESULT 13

Q8TRL3

AC 08TRL3; FRAGMENTARY; PRT, 293 AA

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical protein MALL162.

GN MALL162.

OS Methanosarcina acetivorans.

OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales.

OC Methanosarcinaceae; Methanosarcina.

OX NCBI_TaxID=2214;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN CQA / ATCC 35066 / DSM 2004.

RX MEDLINE=21929760; PubMed=11932238;

PA Galagan J.E., Nishbaum C., Roy A., Endrizzi M.G., MacDonald P.,

PA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,

PA Allen N., Nayler J., Stange-Themann N., Dearlilane F., Johnson P.,

PA Linton J., McEwan P., McKernan K., Tiliavas J., Tirrell A., Ye W.,

PA Zimmer A., Barber P.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,

PA Hedderich P., Ingram-Smith C., Kiermer H.G., Krzycki T.A.,

PA Leigh J.A., Li W., Liu J., Mukhopadhyay R., Reeve J.N., Smith K.,

PA Springer T.A., Tmagam L.A., White O., White P.H., de Macario E.C.,

PA Ferry J.G., Darrell K.F., Ding H., Macario A.J.L., Paulsen I.,

PA Pritchett W., Sowers K.P., Swanson P.V., Zinder S.H., Zinder E.,

PA Metcalf W.W., Birren B.

PT "The genome of Methanosarcina acetivorans reveals extensive metabolic

RT and physiological diversity."

RL Genome Res. 12:542-549(2002)

DR EMBL; AE010783; AAM04583.1;

KW Hypothetical protein; Complete proteome

SQ SEQUENCE 293 AA; 30290 MW; 5420170KFAQ2PPEQ CPO64;

Query Match 100.0%; Score 38, DB 16, Length 293;

Best Local Similarity 66.7%; Pred No. 2 90.02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

Db 168 WSIWHF 173

RESULT 14

Q91065

AC Q91065; PRELIMINARY; PRT, 293 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical protein PA2778.

GN PA2778

OS Escherichia aeruginosa

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

PA Steyer G.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

PA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

PA Garber P.L., Gentry B., Tolentino E., Westbrook-Wadman S., Yuan Y.,

PA Brody L.L., Coulter S.N., Folger K.P., Vas A., Barthig Y., Lim P.M.,

PA Smith K.A., Spencer P.H., Wong G.K.S., Wu Z., Paulsen I.T.,

PA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

PT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen."

RL Nature 406:959-964(2000).

DB EMBL; AE004705; AAC06166.1;

DR InterPro; IPR005074; Peptidase_C39.

DR InterPro; IPR001440; TPR.

DR Pfam; PF03412; Peptidase_C39; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 293 AA; 30016 MW; 574170EFTFTEI W764;

Query Match 100.0%; Score 38, DB 16, Length 293;

Best Local Similarity 66.7%; Pred No. 2 90.02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

Db 101 WPRWHF 106

RESULT 15

Q9EXF1

AC Q9EXF1; PRELIMINARY; PRT, 310 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Putative phosphotransferase P arabidii specific component IIC

DE (Fragment).

GN ATLC.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Listeriaceae; Listeria.

OX NCBI_TaxID=1629;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LO28;

EX MEDLINE=21212979; PubMed=11316371;

PA Saklani-Dusforques H., Fontan E., Goossens P.L.;

PT "Characterisation of a Listeria monocytogenes mutant deficient in D-

RT arabitol fermentation."

RL Res. Microbiol. 152:175-177(2001).

DR EMBL; AC022552; CACC0641.1;

DR InterPro; IPR004703; Gal_spec_IIC.

DR Pfam; PF03611; EIIc-GAT; 1.

KW Transferase.

FT NON_TER 1 1

FT NON_TER 310 310

SQ SEQUENCE 310 AA; 32711 MW; A2P9FE9CPC8ADACF CPO64;

Query Match 100.0%; Score 34, DB 2; Length 310;

Best Local Similarity 66.7%; Pred No. 2 90.02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

Db 18 WNYWHF 23

Search completed, February 20, 2003, 10:43:22

Job time: 25.3333 secs



GenCore version 5.1.3
Copyright 1993 - 2003 Compugen Ltd

QM protein - protein search, using sw model

Run on: February 20, 2003, 10:35:40, Search time 29.333 seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 38

Sequence: 1 WXXWHF 6

Scoring table: EIGSME2EX
Gapop 10 0, Gapext 0 5

Searched: 908470 seqs, 13350620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A Geneseq 101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2004.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2005.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2006.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2007.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2008.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2009.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2010.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the hit's score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	6	21 AAB01492	Peptide which bind
2	38	100.0	6	21 AAB01499	Peptide which bind
3	38	100.0	6	21 AAB01505	Peptide which bind
4	38	100.0	6	21 AAB01506	Peptide which bind
5	38	100.0	6	22 ABR40893	Peptide #8399 enco
6	38	100.0	36	22 AAM61753	Human brain expro
7	38	100.0	36	22 AAM74549	Human bone marrow
8	38	100.0	36	22 AAM74663	Peptide #8700 enco
9	38	100.0	43	22 AAM14683	Novel bone marrow
10	38	100.0	63	23 AAG34992	Human albumin fusl

11	38	100.0	53	23 AAE21233	Human gene 18 enco
12	38	100.0	57	22 AAT66178	Protein bacterium
13	38	100.0	65	23 ABE17322	Human GEF1 protein
14	38	100.0	87	22 AAY65391	Protein bacterium
15	38	100.0	96	22 AAT14777	Novel bone marrow
16	38	100.0	119	23 ABE61339	Human GEF1 protein
17	38	100.0	173	22 AAG80844	Lipid modification
18	38	100.0	173	22 AAT82334	Human muscle skele
19	38	100.0	175	22 ABB93211	Human muscle skele
20	38	100.0	175	22 AAT42371	Human muscle skele
21	38	100.0	256	21 AAG09463	Arabidopsis thalia
22	38	100.0	256	21 AAG44875	Arabidopsis thalia
23	38	100.0	256	21 AAG48293	Arabidopsis thalia
24	38	100.0	258	22 AAG93113	C glutamic prote
25	38	100.0	306	21 AAG65810	Arabidopsis thalia
26	38	100.0	306	21 AAG47194	Arabidopsis thalia
27	38	100.0	324	21 AAG50316	Arabidopsis thalia
28	38	100.0	324	21 AAG44662	Arabidopsis thalia
29	38	100.0	324	21 AAG44664	Arabidopsis thalia
30	38	100.0	325	21 AAG48292	Arabidopsis thalia
31	38	100.0	338	21 AAG58809	Arabidopsis thalia
32	38	100.0	338	21 AAG47193	Arabidopsis thalia
33	38	100.0	376	19 AAG46297	Phytobium meliloti
34	38	100.0	377	19 AAM46298	Brucella abortus D
35	38	100.0	418	15 AAG60501	Lipopolysaccharide
36	38	100.0	423	21 AAG50315	Arabidopsis thalia
37	38	100.0	423	21 AAG44662	Arabidopsis thalia
38	38	100.0	424	22 AAE13431	Coated microsome
39	38	100.0	427	22 AAT20451	Human secreted pro
40	38	100.0	427	22 AAT21939	Novel human neopla
41	38	100.0	448	21 AAG44662	Arabidopsis thalia
42	38	100.0	448	21 AAG44662	Arabidopsis thalia
43	38	100.0	448	21 AAG44662	Arabidopsis thalia
44	38	100.0	452	21 AAG44662	Arabidopsis thalia
45	38	100.0	455	22 AAB35611	Human protein seqs

ALIGNMENTS

RESULT 1
ID AAB01492 standard, peptide, 6 AA.
XX AAB01492;
XX AC AAB01492;
XX DT 09-NEV 2000 (first entry)
XX DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX XX
XX DNA binding, transcription factor, E2F, E2F 1, cell cycle, G1-S;
XX FW activation, transcription, apoptosis, proliferative disorder;
XX KW psoriasis, restenosis.
XX XX
XX OS Synthetic.
XX XX
XX PR WC200044771-A1.
XX XX
XX PD 03-AUG-2000.
XX XX
XX PF 26-JAN 2000, 2000WO 9806227
XX XX
XX PR 26-JAN-1999, 99GB-0001710.
XX XX
XX PA (PROT-) PROTEIN LTD.
XX XX
XX PT Mueller F, Kontermann PE, Montigiani S;
XX XX
XX DR WPI; 2000-532806/48.
XX XX
XX FT Peptides binding to the DNA binding domain of transcription factor E2F
XX and inhibiting cell cycle progression, useful for the treatment of
XX cancer

XX
PS Claim 6; Page 9; 4pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.

XX
SQ Sequence 6 AA;

Query Match 100.0%, Score 38, DB 21; Length 6,
Best Local Similarity 66.7%, Prod No. / Seq. Id.
Matches 4, Conservative 2, Mismatches 0, Indels 0, Gaps 0

QY 1 WXXXWHP 6
|||
DB 1 WVRWHP 6

RESULT 2
AAB01499
ID AAB01499 standard; peptide; 6 AA
AC AAB01499;
DI 08-NOV-2000 (first entry)
DE Peptide which binds to transcription factor E2F 1 DNA binding domain
KW "DNA binding", "transcription factor", "E2F", "E2F 1", "cell cycle", "G1",
KM activating transcription; "proliferation", "apoptosis", "proliferative disorder",
KN "psoriasis", "restenosis".
XX Synthesis:
FH Key Location/Source/Accession
FT Miscellaneous 2 "Any amino acid"
PT Misc difference 3
FT Misc difference 3 /note= "Any amino acid"
XX WO200044771 A1.
PN
XX 03-AUG-2000;
PD 03-AUG-2000;
FI de JAM 2000; JMWMA/JMHWJ?
XX 26-MAY-2000; 03CF 0000010
PR (PROL-) PROLIFIX LTD.
PA Mueller S, Kuntermann SE, Montigiani S;
XX Mueller S, Kuntermann SE, Montigiani S;
DR WI, 2000 512806/48
XX
PT Peptides binding to the DNA binding beta of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Claim 4; Page 9; 42pp; English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.

SQ Sequence 6 AA;

```

Query Match      100.0%, Score 38, DB 21, Length 6;
Best Local Similarity 100.0%, Fred. No. 7.8e+05;
Mismatch 0; Conserved 0; Mismatches 0; Indels 0; Gaps 0;
DB      1 WXXWHF 6
      |||||
      1 WXXWHF 6

RESULT 3
AAB01505
ID      AAB01505 standard; peptide; 6 AA.
XX
AF      AAB01505;
XX
DT      08-NOV-2000 (first entry)
XX
DE      Peptide which binds to transcription factor E2F 1 DNA binding domain.
XX
KW      CNA binding, transcription factor, E2F, E2F 1, cell cycle, DP-1,
KW      activation, transcription, proteins, proliferative disorder;
KW      psoriasis; restenosis.
XX
OS      Synthetic
XX
PN      WO2000044771-A1.
XX
PO      US AUG-2000.
XX
PT      OF TAN 2007; 2006WO 0200027.
XX
PF      20 JAN 2001, 2000-0001710.
XX
PA      (FRODO-) PROLIFIX LTD.
XX
PT      Mueller E, Koberman FE, Montigiani S;
XX
PF      WFI, 2000-00000718.
XX
PT      Peptides binding to the CNA binding domain of transcription factor E2F
PT      and inhibiting cell cycle progression, useful for the treatment of
PT      cancer
XX
XX
XX      Example: Page 26; 42pp, English.
XX
DE      Peptides which bind to the CNA binding domain of transcription
DE      factor E2F and inhibit cell cycle progression may be used as
DE      treatment agents to investigate the interaction between E2F and DP-1,
DE      or the activation of transcription by E2F-1/DP-1 heterodimers. They
DE      may also be used for inhibiting of these and/or cell cycle arrest in
DE      a cell, particularly for treatment of cancer or other proliferative
DE      disorders such as psoriasis and restenosis.
XX
XX
XX      Sequence 6 AA;
XX
XX
XX      Query Match      100.0%, Score 38, DB 21, Length 6;
XX      Best Local Similarity 100.0%, Fred. No. 7.8e+05;
XX      Mismatch 0; Conserved 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      DB      1 WXXWHF 6
XX      |||||
XX      1 WXXWHF 6

RESULT 4
AAB01506
ID      AAB01506 standard; peptide; 6 AA.
XX
AF      AAB01506;
XX
DT      08-NOV-2000 (first entry)
XX

```

DE Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW DNA binding, transcription factor, E2F, E2F-1, cell cycle, DR-1,
KW activation, transcription, apoptosis, proliferative disorder,
KW psoriasis, restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000WO-GR00227.
XX
PR 26-JAN-1999; WAGR-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann PF, Montigiani S;
XX
DR WPI; 2000-532806/48
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
PS Example; Page 26; 42pp, English.
XX
CC Peptides which bind to the DNA binding domain of transcription
CC factor P3F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 38; DR 21; Length 6;
Best Local Similarity 66.7%; Pred. No. 7.8e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:|:|
Db 1 WVAWHF 6

RESULT 5
ABR40893
ID ABB40893 standard; Peptide; 36 AA.
XX
AC ABB40893;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #8399 encoded by human foetal liver single exon probe
XX
KW Human; foetal liver; gene expression, single exon nucleic acid probe
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 27-SEP-2000; 2000US-0236359.
XX

PP 04-OCT-2000; 2000GP-0004063
XX
FA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001 483447/52.
XX
PT Human genome-derived single-exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 33528, 639pp + sequence listing, English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp:wipo.int/pub/published_pat_sequences
XX
SQ Sequence 36 AA;

Query Match 100.0%; Score 38; DR 22; Length 36;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:|:|
Db 19 WRAWHF 24

RESULT 6
AAM61753
ID AAM61753 standard; Protein; 36 AA.
XX
AC AAM61753;
XX
DT 01-APR-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33858.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001 483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX

XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX
OS Homo sapiens.
XX
PN WO200157187-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO US03782.
XX
PR 01-FEB-2000; 2000US-0496914.
PR 20 JUN 2000; 2000US-0594075.
PR 19-JUL-2000; 2000US-0620325.
PR 10-NOV 2000; 2000US-0250683.
XX
PA (HVSF-) HVSFQ INC.
XX
PI Ford JE, Boyle PJ, Tang YF, Liu C, Asanul V, Chao F, Xue AG,
PI Ren F, Drmanac RT;
XX
DR WPI; 2001-488875/53
DR N FTRP; AAE22334
XX
PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX
PS Claim 10; Page 248; 392pp; English.
XX
CC AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or a supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins.
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ Sequence 43 AA;
XX
Query Match 100.0%; Score 39; DB 22; Length 43;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXWHF 6
Db 14 WLIWHF 19
XX
RESULT 10
ABG64892
ID ABG64892 standard; Protein; 53 AA.
XX
AC ABG64892;

XX
DT 27 AUG 2002 (first entry)
XX
DE Human albumin fusion protein #1567.
XX
KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW human serum albumin; HSA; cancer; reproductive disorder;
KW digestive disorder; immune disorder; endocrine disorder;
KW haematopoietic disorder; neural disorder; connective disorder;
KW cytoskeletal; antifertility; antiinflammatory; antiulcer;
KW immunomodulator, anti HIV, antidiabetic, haemostatic, neurotropic;
KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW osteopathic; antiarthritic.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200177137-A1.
XX
PD 18-OCT-2001.
XX
PF 12-APR 2001; 2001WO US11568.
XX
PR 12-APR 2000; 2000US-229358P.
PR 25 APR 2000; 2000US-199384P.
PR 21 FEB 2000; 2000US-256301P
XX
FA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Haseltine WA;
XX
PI WPI; 2002-010886/01.
XX
DR New fusion protein for treating disease e.g. diabetes comprises an
XX albumin fused to a therapeutic protein -
XX
PT Claim 1; Page 1592; 2102pp; English.
XX
CC The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA, also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or
CC disorder that may be modulated by therapeutic protein X. The albumin
CC extends the shelf-life of protein X, and may increase its biological
CC in vitro/in vivo activity. The protein is useful for treating and
CC diagnosing disorders such as cancer, reproductive disorders, digestive
CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
CC (e.g. diabetes), haematopoietic disorders, neural disorders
CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
CC (e.g. osteoporosis, arthritis). AB053326-AB056518 represent albumin
CC fusion proteins of the invention.
XX
SQ Sequence 53 AA;
XX
Query Match 100.0%; Score 38; DB 23; Length 53;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXWHF 6
Db 21 WASWHF 26
XX
RESULT 11
AAE21233
ID AAE21233 standard; Protein; 53 AA.
XX
AC AAE21233;
XX
DT 01 JUL 2002 (first entry)
XX
DE Human gene 18 encoded secreted protein HNNEM45, SEQ ID NO:98.

XX Sequence 57 AA;
SQ Query Match 100.0%; Score 38, PR 22, Length 57;
Best Local Similarity 66.7%; Pred. No. 87;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXWHF 6
Db 4 WFPWHF 9
RESULT 13
ABP07322 standard; Protein; 65 AA.
XX ID ABP07322 standard; Protein; 65 AA.
XX AC ABP07322;
XX DT 24-JUN-2002 (first entry)
XX DE Human GREF protein sequence SEQ ID NO:14626.
XX KW Human; open reading frame; GREF; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29 APR 2000; 2000US-206710P.
XX PA (CTRA-) CTRAGEN CORP.
XX PI Shimkets RA, Leach MD;
XX DR WPI; 2002-106308/14.
XX DR N-PSDB; ABN23074.
XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX PS Disclosure; SEQ ID 14626; 1037pp; English.
XX CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, GREF, where X is 1-11491 (see Table 1
CC in the specification) APT5762 to ABN27252 encode the human GREF
CC proteins given in ABP00010 to ABP11500. GREF proteins are useful for
CC treating or preventing a pathology associated with an GREF-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with GREF associated disorder. GREF polynucleotide
CC sequences can be used in gene therapy. GREF sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. GREF proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N R. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp wipo.int/pub/published_pat_sequences.
XX
XX SQ Sequence 65 AA;
QY Query Match 100.0%; Score 38; PR 23; Length 65;
Best Local Similarity 66.7%; Pred. No. 99;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 27 WKSWHF 32
QY 1 WXXWHF 6
Db 27 WKSWHF 32
RESULT 14
AAU56391 standard; Protein; 87 AA.
XX ID AAU56391 standard; Protein; 87 AA.
XX AC AAU56391;
XX DT 27 FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #17287.
XX KW SAPHO syndrome, synovitis, acne, pustulosis, hyperostosis, osteomyelitis;
KW uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris, enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 29-APR-2001; 2001WO-US12865.
XX PR 21-APR 2000; 2000US-199047P.
XX PR 02 JUN 2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'Esionneuve J, Zhang Y, Jen S, Carter D;
XX DR WPI; 2001-616774/71.
XX DR N-PSDB; AAS59575.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX PS Example 1; SEQ ID No 17586; 1069pp; English.
XX CC Sequences AAU56391-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated cDNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and endophthalmitis.
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC immunoglobulin expression and activity. The antibodies may also be used as
CC reagents to test for disease infections. The antibodies may also be used as
CC diagnostic agents for identifying bone marrow disease, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Nucleotide sequence: 1715 (1113-1113) (1113-1113) (1113-1113)
CC specification, but was obtained in electron's format directly from WIPD
CC at ftp.wipo.int/pub/published/pub_sequences.
XX
SQ Sequence 37 AA,
Query Match: 100.00, Score 35, SE 12, Length 37,
Best Local Similarity: 66.78; Pred No: 1.3e+02;
Matches 4, Conservative 2, Mismatches 0, Indels 0, Gaps 0,
Db 34 WFPWHP 29

RESULT 15
AAU14777

ID AAU14777 standard; Protein; 96 AA.

XX AAU14777;

DT 24-OCT-2001 (first entry)

DE Novel bone marrow polypeptide #176.

KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.

OS Homo sapiens.

PN M0200157187-A2.

PD 09-AUG 2001.

PF 05-FEB-2001; 2001WD-US03/82.

PR 03-FEB-2000; 2000US-0496914.

PR 20-JUN 2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 30-NOV 2000; 2000US-0250683.

PA (NYSE-) HYSEQ INC.

PI Foid JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou F, Xue AJ;
PI Ren F, Dymnac RT;

DR MPI; 2001-48875/53.
DR N-PSDB; AAS23082.

PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy.

PS Claim 10; Page 137; 392pp; English.

CC AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequence may also be used as DNA
CC probes in diagnostic assays to detect and quantify the presence of
CC similar nucleic acid sequences in samples, and therefore which patients

CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify, monitor, or of bone regression and activity.
CC The anti bone marrow protein antibodies and antigens may also be used
CC to detect regulatory expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g., by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoietic activity, and consequently in the
CC treatment of myeloid or lymph cell disorders, in tissue regeneration,
CC such as wound healing, as a nutritional supplement, and in treatment of
CC severe combined immunodeficiency (SCID).

XX Sequence 36 AA;

Query Match: 100.00, Score 32, SE 12, Length 36,
Best Local Similarity: 66.78; Pred No: 1.4e+02;
Matches 4, Conservative 2, Mismatches 0, Indels 0, Gaps 0,
Db 67 WFPWHP 72

OY 1 WXXWHP 6

DB 67 WFPWHP 72

Search completed: February 20, 2003, 10:47:37
Job time : 30.3333 secs

GenCore version 5.1.3
Copyright 1993 2002 GenProgen Ltd

OM protein - protein search, using sw model

File: us-09-912-414-11 Search time: 12 seconds
(without alignments)
48.067 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 6
Sequence: 1 WXXWHF 6

Scoring table: OLIGOX
Gapop 60.0, Gapext 60.0

Searched: 293224 seqs, 96134422 residues

Word size: 0

Total number of hits satisfying chosen parameters: 293224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post processing: listing first 45 summaries

Database: PIR_73:*

1: pirl:1:
2: pirl:2:
3: pirl:3:
4: pirl:4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	6	100.0	230	D70847	hypothetical prote
2	6	100.0	241	A71020	hypothetical prote
3	6	100.0	260	P71393	probable D,P carb
4	6	100.0	293	A83299	hypothetical prote
5	6	100.0	358	S43876	site-specific DNA
6	6	100.0	368	A87296	modification methy
7	6	100.0	386	A97456	adenine-specific m
8	6	100.0	394	AL2674	adenine DNA methyl
9	6	100.0	403	D90105	putative SAR DNA-b
10	6	100.0	403	AF3432	site-specific DNA-
11	6	100.0	423	AH1407	PTS system galacti
12	6	100.0	423	AH1783	PTS system galacti
13	6	100.0	424	JC5891	omega 6 desaturase
14	6	100.0	424	T07742	omega-6 desaturase
15	6	100.0	439	T01807	hypothetical prote
16	6	100.0	447	T08137	probable omega-6 d
17	6	100.0	447	T17299	hypothetical prote
18	6	100.0	448	D85362	hypothetical prote
19	6	100.0	452	H01172	probable PTS syste
20	6	100.0	452	AH0964	probable PTS syste
21	6	100.0	462	H96018	hypothetical prote
22	6	100.0	472	AG2417	hypothetical prote
23	6	100.0	475	B84171	hypothetical prote
24	6	100.0	487	T32941	hypothetical prote
25	6	100.0	489	P89549	hypothetical prote
26	6	100.0	497	AH2079	alpha-amylase (imp
27	6	100.0	497	T40586	nucleolar protein
28	6	100.0	504	S48650	hypothetical prote
29	6	100.0	508	G90372	4-hydroxyphenylase

30	6	100.0	508	2	T50180	nucleolar protein
31	6	100.0	511	2	S59322	nucleolar protein
32	6	100.0	520	2	P06602	nucleolar protein
33	6	100.0	550	2	T06379	SAR DNA-binding pr
34	6	100.0	560	2	T06377	SAR DNA-binding pr
35	6	100.0	640	2	T06671	Y4-9 protein Rhi
36	6	100.0	718	2	T29448	hypothetical prote
37	6	100.0	774	2	A71379	hypothetical prote
38	5	83.3	40	2	F81511	hypothetical prote
39	5	83.3	41	2	D82458	hypothetical prote
40	5	83.3	42	2	S02108	ethylene responsiv
41	5	83.3	47	2	T04365	hypothetical prote
42	5	83.3	52	2	A07009	hypothetical prote
43	5	83.3	52	2	F07980	hypothetical prote
44	5	83.3	53	2	F89871	hypothetical prote
45	5	83.3	54	2	G90680	hypothetical prote

ALIGNMENTS

RESULT 1

D70847

hypothetical protein RV0059 - Mycobacterium tuberculosis (strain H37PV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 1 / Jul-1998 #ext_change 20-Jun-2000

C/Accession: D70847

R/Cole, S.T.; Brosch, R.; Falkow, S.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroft, S.; Rajandream, M.A.; Rogers, D.; Rupp, S.; Seeger, R.; Skellern, S.; Squires, S.; Nature 393, 537-544, 1998

A/Authors: Squires, R.; Sulston, D.E.; Taylor, K.; Whitehead, S.; Barrett, R.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500, with 9896987, PMID:974230

A/Accession: D70847

A/Status: preliminary, nucleic acid sequence not shown, translation not shown

A/Molecule type: DNA

A/Residues: 1-230 <2DB>

A/Cross-references: 58 AL031428, 58 AL103456, NID 3106154, F000561040.1, P1004280710

A/Experimental source: strain H37PV

C/Genetics:

A/Genes: RV0059

C/Superfamily: Mycobacterium tuberculosis hypothetical protein RV0059

Query Match

Best Local Similarity 66.78% Score 6, Gap 0, Length 230

Matches 4, Conservative 0, Mismatches 0, Indels 0, Gaps 0

QY 1 WXXWHF 6

DE 26 WIVWHF 31

RESULT 2

A71020

hypothetical protein PH1454 - Pyrococcus horikoshii

C/Species: Pyrococcus horikoshii

C/Date: 14-Aug-1998 #sequence_revision 14 Aug 1998 #ext_change 20-Jun-2000

C/Accession: A71020

R/Kawarabayashi, Y.; Sawada, M.; Hattawa, H.; Hattawa, Y.; Hino, Y.; Yamashita, S.; Sekine, M.; Ohnuki, Y.; Sunahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, T.; Kusida, N.; Oguchi, A. Res. 5, 55-76, 1998

A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archa

A/Reference number: A71000, MIMD:08344137, PMID:9670134

A/Accession: A71020

A/Status: preliminary, nucleic acid sequence not shown, translation not shown

A/Molecule type: DNA

A/Residues: 1-241 <2XAM>

A/Cross-references: 58 AF000607, NID 3106153, F00056104.1, P1004280710

A/Experimental source: strain OT3

A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C/Genetics:

A/Genes: PH1454

Query Match: 100.0% Score 6: DB 1: Length 249
Best Local Similarity: 66.7% Prod No. 99
Matches: 4, Conservative: 2, Mismatches: 0, Indels: 0, Gaps: 0

QY 1 WXXWHF 6
DB 193 WQIWHF 198

RESULT 3

Species: *Syphilis spirochete*
Accession: B11353
Title: Complete genome sequence of *Syphilis spirochete*, the syphilis spirochete.
Reference number: A11250; MIMD:9833270; PMID:9665876
Accession: B11353
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-260 <COB>
Cross-references: SE:AC010001, MIM:000001, MIM:000001, MIM:000001
Experimental source: strain N10015
Genetics: A:Gene: TP0221

Query Match: 100.0% Score 6: DB 1: Length 249
Best Local Similarity: 66.7% Prod No. 99
Matches: 4, Conservative: 2, Mismatches: 0, Indels: 0, Gaps: 0

QY 1 WXXWHF 6
DB 220 WEPWHF 225

RESULT 4

Species: *Pseudomonas aeruginosa* (strain PA01)
Accession: A83299
Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.
Reference number: A83299; MIMD:9833270; PMID:9665876
Accession: A83299
Status: preliminary
Molecule type: DNA
Residues: 1-293 <STO>
Cross-references: SE:AC010001, MIM:000001, MIM:000001, MIM:000001
Experimental source: strain PA01
Genetics: A:Gene: PA2778

Query Match: 100.0% Score 6: DB 1: Length 249
Best Local Similarity: 66.7% Prod No. 99
Matches: 4, Conservative: 2, Mismatches: 0, Indels: 0, Gaps: 0

QY 1 WXXWHF 6
DB 101 WPRWHF 106

RESULT 5

Species: *Caulobacter crescentus*
Accession: A97456
Title: Complete genome sequence of *Caulobacter crescentus*.
Reference number: A97456; MIMD:9833270; PMID:9665876
Accession: A97456
Status: preliminary
Molecule type: DNA
Residues: 1-358 <ZMB>
Cross-references: GB:AE007869; PID:AKR6602.1; PID:GL5155772; GSPR:GN00169
Genetics: A:Gene: AGR_C_1453

Species: *Caulobacter crescentus*

Accession: A97456
Title: Complete genome sequence of *Caulobacter crescentus*.
Reference number: A97456; MIMD:9833270; PMID:9665876
Accession: A97456
Status: preliminary
Molecule type: DNA
Residues: 1-358 <ZMB>
Cross-references: GB:AE007869; PID:AKR6602.1; PID:GL5155772; GSPR:GN00169
Genetics: A:Gene: AGR_C_1453

Query Match: 100.0% Score 6: DB 1: Length 358
Best Local Similarity: 66.7% Prod No. 99
Matches: 4, Conservative: 2, Mismatches: 0, Indels: 0, Gaps: 0

QY 1 WXXWHF 6
DB 332 WTYWHF 337

RESULT 6

Species: *Caulobacter crescentus*
Accession: A97296
Title: Complete genome sequence of *Caulobacter crescentus*.
Reference number: A97296; MIMD:9833270; PMID:9665876
Accession: A97296
Status: preliminary
Molecule type: DNA
Residues: 1-358 <STO>
Cross-references: GB:AE007869; PID:AKR6602.1; PID:GL5155772; GSPR:GN00169
Genetics: A:Gene: AGR_C_1453

Query Match: 100.0% Score 6: DB 1: Length 358
Best Local Similarity: 66.7% Prod No. 99
Matches: 4, Conservative: 2, Mismatches: 0, Indels: 0, Gaps: 0

QY 1 WXXWHF 6
DB 332 WTYWHF 337

RESULT 7

Species: *Caulobacter crescentus*
Accession: A97456
Title: Complete genome sequence of *Caulobacter crescentus*.
Reference number: A97456; MIMD:9833270; PMID:9665876
Accession: A97456
Status: preliminary
Molecule type: DNA
Residues: 1-358 <ZMB>
Cross-references: GB:AE007869; PID:AKR6602.1; PID:GL5155772; GSPR:GN00169
Genetics: A:Gene: AGR_C_1453

C:Superfamily: site-specific DNA-methyltransferase (adenine-specific) HpaI

Query Match 100.0%; Score 6; DB 2; Length 386;

Best Local Similarity 66.7%; Pred. No. 1.4e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

DB 357 WTFWHF 362

RESULT 9

AD2674

adenine DNA methyltransferase (imported) Agrobacterium tumefaciens (strain C58, Dupont

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11 Jan-2002 #text_change 17-May-2002

C:Accession: AD2674

R:Wood, D.W.; Setubal, J.C.; Yarb, P.; Marks, P.; Chen, D.; Wood, G.F.; Chen, Y.; Wood, L.

et al.; G. J. Giller, W. J. Grant, G. J. Gentile, D. J. Kutyavin, T. J. Levy, R. J. Li, M. J. McClellan

J. Karp, P. J. Romero, P. J. Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tan, Y.; Riddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon Kamm,

Steff, E. W.

A:Title: The genome of the natural genetic engineering agent Agrobacterium tumefaciens C58

A:Reference number: AB2577; PMID:11743193

A:Accession: AB2577

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-386 <KUR>

A:Cross-references: GR:AB006669; EMBL:AF141810; FIP:31773416; SBFTE:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0794

A:Map position: circular chromosome

C:Superfamily: site-specific DNA-methyltransferase (adenine-specific) HpaI

Query Match 100.0%; Score 6; DB 2; Length 106;

Best Local Similarity 66.7%; Pred. No. 1.4e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

DB 357 WTFWHF 362

RESULT 9

D90105

putative GAP DNA binding protein 1 (imported) - Galliardia theta nucleomorph

C:Species: nucleomorph Galliardia theta

A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: D90105

R:Donaghy, S.; Zenger, S.; Frankham, M.; Pearson, M.; Penny, S.; Terry, L. T.; Wu, X.; Felt

Nature 410, 1091-1096, 2001

A:Title: The highly reduced genome of an enslaved algal nucleus.

A:Reference number: AA0083; PMID:11233671; PMID:11233671

A:Accession: D90105

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <DOU>

A:Cross-references: GR:AB015510; EMBL:AF152677; FIP:31773416; SBFTE:GN00186

C:Genetics:

A:Map position: 2

A:Genome: nucleomorph

C:Keywords: nucleomorph

Query Match 100.0%; Score 6; DB 2; Length 394;

Best Local Similarity 66.7%; Pred. No. 1.4e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

DB 169 WYSWHF 174

RESULT 10

AF3432

site-specific DNA methyltransferase (adenine specific) (EC 2.1.1.72) (imported) Brucella

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01 Feb-2002 #text_change 17-May-2002

C:Accession: AF3432

R:DeVechio, V.G.; Kapatali, V.; Redkar, R.; Pata, S.; Majer, G.; Lee, T.; Ivanova, N.

et al.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzei, F.H.; Haggis, S.; Goughlan, D.; Lohr, S.

Proc Natl Acad Sci USA 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3850; PMID:11766698

A:Accession: AF3432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-403 <KUR>

A:Cross-references: GR:AB006017; EMBL:AF152677; FIP:31773416; SBFTE:GN00186

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME11444

A:Map position: 1

C:Superfamily: site-specific DNA methyltransferase (adenine-specific) HpaI

C:Keywords: methyltransferase, 3-adenosylmethionine

Query Match

100.0%; Score 6; DB 2; Length 403;

Best Local Similarity 66.7%; Pred. No. 1.5e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

DB 373 WTFWHF 378

RESULT 11

AH1407

PTS system galactitol specific enzyme IIC component homolog 1 (imported) - Listeria

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27 Nov 2001 #text_change 14 Dec 2001

C:Accession: AH1407

R:Glaser, P.; Frangoul, D.; Buchrieser, G.; Amend, A.; Bader, F.; Berthel, J.; Fischler, A.

et al.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Ertalan, K.D.; Fritsch, H.

et al.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kretz, T.; Kohn, W.; Kurst, F.; Korth, G.; Madec, F.; Maillochon, A.; Mat

ok, C.; Schuster, T.; Stess, R.; Teyssie, A.; Vazquez-Poland, P.A.; Weiss, H.; Weiland, H.

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; PMID:11676669

A:Accession: AH1407

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-423 <GLA>

A:Cross-references: GR:AB006010; EMBL:AF152677; FIP:31773416; SBFTE:GN00186

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo2665

C:Superfamily: phosphotransferase enzyme II galactitol specific

Query Match 100.0%; Score 6; DB 2; Length 423;

Best Local Similarity 66.7%; Pred. No. 1.5e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

DB 126 WNYWHF 131

RESULT 12

AH1783

PTS system galactitol specific enzyme IIC component homolog 1 (imported) - Listeria

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27 Nov 2001 #text_change 14-Dec-2001

C:Accession: AH1783

GenCore version 5.1.3
Copyright (c) 1991 - 2003 Copyright Ltd

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:00:17, Search time 11.3333 seconds
(without alignments)
21.958 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 6
Sequence: 1 WXXWHF 6

Scoring table: OligoDX
Gapop 60.0, Gapext 60.0

Searched: 112832 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	6	100.0	358	1 MT01_CAU0P	Q45971 Caulobacter
2	6	100.0	376	1 MT01_RHIME	Q30664 rhizobium m
3	6	100.0	377	1 MT01_RPHAR	Q30670 bruceella ab
4	6	100.0	424	1 P06C1_S0YBN	P48628 glycine max
5	6	100.0	426	1 V412_APAAT	Q44568 arabidopsis
6	6	100.0	443	1 P06C1_BRAMA	P48627 brassica na
7	6	100.0	448	1 P06C1_APAAT	P48612 arabidopsis
8	6	100.0	479	1 CATA_PSEPT	Q59714 pseudomonas
9	6	100.0	504	1 SIK1_YEAST	Q12460 saccharomyc
10	6	100.0	511	1 N0P5_YEAST	Q12499 saccharomyc
11	6	100.0	529	1 N0P5_YEAST	Q12499 homo sapien
12	6	100.0	544	1 N0P5_YEAST	Q12499 ratius norv
13	6	100.0	646	1 Y420_RHISM	P55386 rhizobium s
14	6	100.0	718	1 P1SR_CAEER	Q32049 caenorhabdi
15	5	83.3	82	1 TXA1_ACTED	Q60372 actinia eys
16	5	83.3	85	1 VG73_BPMIL5	Q65288 mycobacteri
17	5	83.3	95	1 ESAC_MYCIB	Q65288 mycobacteri
18	5	83.3	97	1 VC49_CVACA	Q67418 cyanidium c
19	5	83.3	101	1 ATPK_YEAST	Q06405 saccharomyc
20	5	83.3	103	1 ES6D_MYCTU	Q05440 mycobacteri
21	5	83.3	104	1 Y491_SYNY3	Q67420 syne-bactys
22	5	83.3	104	1 Y491_YEAST	P47080 saccharomyc
23	5	83.3	105	1 N1GM_HUMAN	Q35178 homo sapien
24	5	83.3	108	1 N1GM_P0VIN	Q22374 hcs +3979
25	5	83.3	111	1 VPX_HV2ER	Q74122 human immu
26	5	83.3	111	1 YFBW_SALTI	P81891 salmonella
27	5	83.3	111	1 YFBW_SALTY	O52328 salmonella
28	5	83.3	112	1 TX1A_AGRAP	P15969 agelenopsis
29	5	83.3	112	1 VPX_HV2CA	P24110 human immu
30	5	83.3	112	1 VPX_HV2D1	P17740 human immu
31	5	83.3	112	1 VPX_HV2G1	P18045 human immu
32	5	83.3	112	1 VPX_HV2N2	P05915 human immu
33	5	83.3	112	1 VPX_HV2P0	P06629 human immu

34	5	83.3	112	1 VPX_HV2SB	P12454 human immu
35	5	83.3	112	1 VPX_HV2ST	P20881 human immu
36	5	83.3	112	1 VPX_HV2M1	P05917 simian immu
37	5	83.3	112	1 VPX_HV2MK	P05916 simian immu
38	5	83.3	112	1 VPX_HV2ML	P11266 simian immu
39	5	83.3	112	1 VPX_HV2S4	P12514 simian immu
40	5	83.3	112	1 VPX_HV2SP	P19508 simian immu
41	5	83.3	112	1 VPX_HV2BE	P19509 human immu
42	5	83.3	118	1 PA21_NAJME	P00509 naja melano
43	5	83.3	118	1 PA21_NAJMO	P00502 naja mossam
44	5	83.3	118	1 PA22_NAJMO	P00502 naja mossam
45	5	83.3	118	1 PA23_NAJMO	P00504 naja mossam

ALIGNMENTS

RESULT 1

MT01_CAU0P STANDARD; PRT; 358 AA.

ID MT01_CAU0P

AC Q45971; 30-MAY-2000 (Rel. 39, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Modification methylase Ccrm1 (EC 2.1.1.72) (Adenine-specific

DE methyltransferase Ccrm1) (M.Ccrm1).

GN CCRM1M OR CCRM OR CCR0378.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CB15N / NA1000;

FX MEDLINE=94118303; PubMed=8289276;

RA Zweiger G, Marczyński G, Shapiro L;

RT "A Caulobacter DNA methyltransferase that functions only in the

RT predivisional cell."

RL J. Mol. Biol. 235:472-485(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CP15;

FX MEDLINE 2173628, PubMed=1259647;

RA Nieman W.C., Feldberg J.F., Laut M.C., Falsen T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alvey M.P.P., Chitt W., Willy J.S.,

RA Potocka J., Nelson W.C., Newton A., Stephens C., Phadke N.D., Fry B.,

RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

PA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Porey K.,

PA Utterback T., Tran P., Wolf A., Vamathevan J., Emolaeva M., White O.,

PA Salzberg S.L., Venter D.C., Shapiro L., Fraser C.M.;

PT "Complete genome sequence of Caulobacter crescentus."

FI Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

CC FUNCTION: THIS METHYLASE REGUENZES THE DOUBLE STRANDED SEQUENCE

CC GATC AND CAUSES SPECIFIC METHYLATION ON A 2 ON BOTH STRANDS.

CC APPARAS TO CONTRIBUTE TO THE ADAPTATE CELL CYCLE CONTROL OF DNA

CC REPLICATION AND CELLULAR MORPHOLOGY.

CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-

CC adenosyl-L-homocysteine + DNA 6-methyladenine.

CC This Swiss Prot entry is copyrighted by the EMBL collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announc/>

CC or send an email to license@sib-sib.ch).

CC EMBL, U01032, AAA18913.1, -

DR EMBL; A0005711; AAK22365.1; -

DR HSSP; P11409; 1BOO.

DR FEPASE; 2535; M.CCRM1.

DR TIGR; G00378; -

[illegible]

RESULT 2	
TEST	REFERENCE
10	TEST REFERENCE
	STANDARD
	FRT
	376 AA

30-MAY-2000 (Ref: 39, Created)
 30-MAY-2000 (Ref: 39, Last sequence update)
 15-JUN-2002 (Ref: 41, Last annotation update)
 DE Modification methylase SmelF (Ref: 2,1,1,72) (Adenine-specific
 methytransferase SmelF) (Mismeijer, McCormick).
 SM strain of *Escherichia coli* O157:H7.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC Bacteria; Proteobacteria, Alpha-Proteobacteria, Rhizobiales
 CC Rhizobiaceae; Sinorhizobiaceae.
 OX NCBI_TaxID 382,
 RN 1,
 RF SEQUENCE FROM NIA.
 RC STRAIN-1021;
 RX MEDLINE 3744013, PubMed 3224447,
 RA Wright R, Stephens C, Shapiro B,
 RT "The *CaM* two catalytic domains in *CaM* are essential for the
 RT of proteobacteria, and its essential functions are conserved in
 RT Rhizobium meliloti and Caulobacter crescentum",
 RT C. Bacteriol. 179:5869-5877(1997).
 RN 12,
 RT SEQUENCE FROM NIA.
 RC STRAIN-1021;
 RX MEDLINE 333633, PubMed 1444444,
 RA Capela A, Bally-Balazs A, Eddy M, Gillis E, Tison E, et al.,
 RA Bolstad J, Beckers A, Eddy M, Gillis E, Tison E, et al.,
 RA "The *CaM* two catalytic domains in *CaM* are essential for the
 RA of proteobacteria, and its essential functions are conserved in
 RA Rhizobium meliloti and Caulobacter crescentum",
 RA C. Bacteriol. 179:5869-5877(1997).
 RT "Analysis of the amino acid sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021".
 RL Proc. Natl. Acad. Sci. USA 96:6977-6982(1999).
 CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE NOBLE STANDED SEQUENCE
 CC GANTC AND CAUSES SITE-SPECIFIC METHYLATION ON A 3' OH BOTH STANDED
 CC CCM-MEDIATED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
 CC ATTENDS TO CONTRIBUTE TO THE ACQUISITION OF THE NATURE OF THE
 CC REPLICATION AND CELLULAR MORPHOLOGY.
 CC -2- CATALYTIC ACTIVITY: 2-adenosine-methyltransferase, DNA adenine-3
 CC adenine-3-methyltransferase, DNA adenine-3-methyltransferase.
 CC THIS ENTRY FROM NIA IS AVAILABLE TO THE PUBLIC THROUGH A PUBLIC
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTANDING
 CC THE European Bioinformatics Institute. There are no restrictions on
 CC use by any third party. However, if you are using the data for
 CC modified and this statement is not removed, may be by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk)

DR	EMBL; AF01694; AAB71350.1; .
DE	EMBL; AL831795; CAC45498.1; .
DF	REFSEQ; J264; M.Smeip.
DR	InterPro; IP001091; CN4_Mettransf.
DR	InterPro; IP002205; DINO_mifase.
DR	InterPro; IP002204; NC/N4_Mtase.
DE	InterPro; IP000000; No_Mtase.
DR	InterPro; IP000051; SAM_bind.
DE	IJUN; EF155; NC/N4_Mtase; 1.
DE	PRINTS; PR00006; DINOCTRACE.
DE	PRINTS; PR00009; EOLINCTRACE.
DE	IP_0010; IPR000000; No_Mtase; 1.
KW	Transferrase; Methyltransferase; DNA replication; Complete proteome.
FT	COILED1 135 141 NMTNTR [C] CDADDO (IN REF. 1).
FT	CONFICT 157 157 F [A] (IN REF. 1).
FT	SELENCE 324 AA, 4148 MW, 10047800 Da, A-3264;

Query Match: 100%, Score 5, IP 1, Length 375;
Best Local Similarity: 66.7%; Pred. No. 42;
Matches: 4, Conservative: 2, Mismatched: 0, Indels: 0, Gaps: 0

QY	1	WXXWHE	6
	:	:	:
	:	:	:
Db	347	WTFWHF	352

```

RESULT 3
MTBI_BPRUB
ID MTBI_BPRUB STANDARD: FRT 377 AA.
AC 030670;
DI 30-MAY-2000 (Rel. 35, last sequence update)
DI 01-MAY-2000 (Rel. 30, last sequence update)
DI 16-OCT-2001 (Rel. 40, last annotation update)
DE Modification methylase BabI (M.BabI) (M.Corm.)
DE methyltransferase BabI (M.BabI) (M.Corm.).
GN BABI OR CORM.
OS Brucella abortus.
OC (act-14) Proteobacteria, alpha subdivision, Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID 235;
RN [1]
RT SEQUENCE FROM N.A.
RC STRAIN-02308;
EX REFERENCE 3741033, PubMed 2204147;
RA Wright P., Stephens C., Shapiro L.;
RT "The Corm DNA methyltransferase is widespread in the alpha subdivision
of proteobacteria, and its essential functions are conserved in
Rhizobium meliloti and Caulobacter crescentus."
RT J. Bacteriol. 179:665-671 (1997).
OC FUNCTION: THIS MEMBRANE PROTEIN AND THE PROTEIN-TRANSFER OFFICER
HARP AND HARP-TRANSFER METHYLATION ON A 2 ON EACH STRAND.
OC WHEN RELEASED METHYLATION HAS IMPORTANT CELLULAR FUNCTIONS.
OC APPEARS TO CONTRIBUTE TO THE ACCURATE CELL CYCLE CONTROL OF DNA
OC REPLICATION AND CELLULAR MORPHOLOGY.
OC CATALYTIC ACTIVITY: 2-adenosyl-5-methylthio-3'-DNA-adenine-5'-S
adenosyl-5-homocysteine + DNA-6-methyladenosine.
CC THIS SWISS PROT entry is annotated to be produced through a translation
between the Swiss Institute of Bioinformatics and the EMBL outstation. It
the EMBL outstation facilities. There are no restrictions on its
use by any party. Information and a link to the content is in any way
classified and this statement is to be removed from any and for commercial
purposes. A license agreement (see http://www.isb.ch/announce/)
is good in order to be available (lib.ch).
CC EMBL, AF011907, AB013511; -.
CC EMBASE, 3263, M.BabI.
CC FURTHER INFORMATION: CM4 Methylase.
CC FURTHER INFORMATION: BabI Methylase.
CC FURTHER INFORMATION: BabI Methylase.
CC INTERPRO: IPR002052, N6_Mtase.
CC INTERPRO: IPR000051, SAM_bind.
CC Pfam: PF01555, N6_N4_Mtase; 1.

```

DR PRINTS; PR00506; D21N6MTFRASE.
DR PRINTS; PR00508; S21N4MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1
KM Transferase; Methyltransferase; DNA replication.
SQ SPROINGF 377 AA; 42200 MW; 667049A05580839D CRC64;

Query Match 100.0%; Score 6; DB 1; Length 377;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
DB 347 WTFWHF 352

RESULT 4

FDR6 SOYBN

ID FDR6 SOYBN STANDARD; PRT; 424 AA.
AC P48628;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Omega-6 fatty acid desaturase, chloroplast precursor (F011499)

OS Glycine max (Soybean)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons; Rosales;

OC Eurosidia I; Fabiales; Fabaceae; Papilionoideae; Phaseolidae; Glycine

OX NCBI_TaxID=3847;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Seed;

RX MEDLINE=94345008; Pubmed 8066133;

RA Hitz W.D., Carlson T.J., Booth J.P. Jr., Kinney A.D., Stecca K.L.,

RA Yadav N.S.;

RT "Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA

RT and its expression in a cyanobacterium.";

RL Plant Physiol 105:635-641(1994)

CC - FUNCTION: CHLOROPLAST OMEGA-6 FATTY ACID DESATURASE INTRODUCES

CC THE SECOND DOUBLEBOND IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY

CC ACIDS. IMPORTANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT

CC TO USE FERREDOXIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS

CC ESTERIFIED TO GLYCEROLIPIDS. SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.

CC - PATHWAY: Polyunsaturated fatty acid biosynthesis.

CC - SUBCELLULAR LOCATION: CHLOROPLAST; MEMBRANE BOUND (PROBABLE)

CC - DOMAIN: THE HISTIDINE BOX DOMAIN MAY CONTAIN THE ACTIVE SITE

CC AND/ OR BE INVOLVED IN METAL ION BINDING.

CC - SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch)

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

QY 1 WXXWHF 6
DB 251 WXXWHF 256

RESULT 5
Y412 ARATH
ID Y412 ARATH STANDARD; PRT; 439 AA.
AC 004658;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein A5927120

GN A5927120 OR TM021804.12 OR T21B4_30.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons; Rosales;

OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, Columbia;

RX MEDLINE=21016721; Pubmed 11130714;

RA Tabata S., Yaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,

RA Miyajima N., Sasamori S., Kimura T., Hoshino T., Kawashima K.,

RA Kohata M., Watanabe M., Matsuno A., Muraki A., Nakayama S.,

RA Nishikawa M., Haruo K., Okumura S., Shirai S., Takeuchi C., Wada T.,

RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,

RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston P.,

RA Haberman P., Murray J., Johnson P., F-Hling T., Nelson P.,

RA Stoneking T., Pepin K., Spieth J., Sakon M., Armstrong J., Becker M.,

RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dente M.,

RA Du H., Edwards J., Fryman J., Haakensen E., Lamar E., Latreille P.,

RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromwater C.,

RA Wagner-Mederson C., William A., Yeakum M., Bell M., Dedhia N.,

RA Parnell L., Shah P., Rodriguez M., Hoon Sae I., Vill P., Baker D.,

RA Kirchhoff K., Toth K., King L., Rahner A., Miller P., Maria M.,

RA Wollensack F., Wollensack W.F., Wilson P.K., Murphy G., Hancock J.,

RA Volckaert G., Wambolt P., Duestelheide A., Stiekema W., Fohl T.,

RA Eickel P., Taylor N., Hartley N., Port F., Jones A.S.,

RA Langham S.A., McCullagh E., Fokken G., Orymoprez B., Zimmermann W.,

RA Ramsberger U., Wedler H., Balke K., Wedler E., Peters S.,

RA van Staveren M., Dikse W., Moolman P., Klein Lankhorst P.,

RA Weitzenecker T., Bothe G., Rose M., Hauf J., Reinleiser S., Hempel S.,

RA Feldmann M., Lambeth S., Villarejo R., Gielen J., Ardiles W.,

RA Bente G., Lemcke K., Kolesov G., Mayer K., Fuhl S., Scholl H.,

RA Schueller C., Gascaria P., Mewes H.W., Benan M., Franke F.,

RT "Sequence and analysis of chromosome 6 of the plant Arabidopsis

RT thaliana.";

RL Nature 409:823-826(2000).

CC - SIMILARITY: BELONGS TO THE NDR/NDP56 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch)

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

[illegible]

DE Omega 6 fatty acid desaturase, chloroplast precursor (F01.1.14.99).
GN FAD6 OR FAD6 OR A14G30950 OR F6118.140.
OS Arabidopsis thaliana (Mouse ear cress).
OC Eudicotyledons, Eudicotyledons, Eudicotyledons, Core eudicots, Rosidae
OC Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae
OC Equisetids II, Brassicales, Brassicaceae, Arabidopsis.
OX NCBI_Taxid:3702;
RN
RP
RT
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
RP
RS
RT
RU
RV
RW
RX
RY
RZ
S0
S1
S2
S3
S4
S5
S6
S7
S8
S9
T0
T1
T2
T3
T4
T5
T6
T7
T8
T9
U0
U1
U2
U3
U4
U5
U6
U7
U8
U9
V0
V1
V2
V3
V4
V5
V6
V7
V8
V9
W0
W1
W2
W3
W4
W5
W6
W7
W8
W9
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
Y0
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
Y9
Z0
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Z9
[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]
[49]
[50]
[51]
[52]
[53]
[54]
[55]
[56]
[57]
[58]
[59]
[60]
[61]
[62]
[63]
[64]
[65]
[66]
[67]
[68]
[69]
[70]
[71]
[72]
[73]
[74]
[75]
[76]
[77]
[78]
[79]
[80]
[81]
[82]
[83]
[84]
[85]
[86]
[87]
[88]
[89]
[90]
[91]
[92]
[93]
[94]
[95]
[96]
[97]
[98]
[99]
[100]
[101]
[102]
[103]
[104]
[105]
[106]
[107]
[108]
[109]
[110]
[111]
[112]
[113]
[114]
[115]
[116]
[117]
[118]
[119]
[120]
[121]
[122]
[123]
[124]
[125]
[126]
[127]
[128]
[129]
[130]
[131]
[132]
[133]
[134]
[135]
[136]
[137]
[138]
[139]
[140]
[141]
[142]
[143]
[144]
[145]
[146]
[147]
[148]
[149]
[150]
[151]
[152]
[153]
[154]
[155]
[156]
[157]
[158]
[159]
[160]
[161]
[162]
[163]
[164]
[165]
[166]
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177]
[178]
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194]
[195]
[196]
[197]
[198]
[199]
[200]
[201]
[202]
[203]
[204]
[205]
[206]
[207]
[208]
[209]
[210]
[211]
[212]
[213]
[214]
[215]
[216]
[217]
[218]
[219]
[220]
[221]
[222]
[223]
[224]
[225]
[226]
[227]
[228]
[229]
[230]
[231]
[232]
[233]
[234]
[235]
[236]
[237]
[238]
[239]
[240]
[241]
[242]
[243]
[244]
[245]
[246]
[247]
[248]
[249]
[250]
[251]
[252]
[253]
[254]
[255]
[256]
[257]
[258]
[259]
[260]
[261]
[262]
[263]
[264]
[265]
[266]
[267]
[268]
[269]
[270]
[271]
[272]
[273]
[274]
[275]
[276]
[277]
[278]
[279]
[280]
[281]
[282]
[283]
[284]
[285]
[286]
[287]
[288]
[289]
[290]
[291]
[292]
[293]
[294]
[295]
[296]
[297]
[298]
[299]
[300]
[301]
[302]
[303]
[304]
[305]
[306]
[307]
[308]
[309]
[310]
[311]
[312]
[313]
[314]
[315]
[316]
[317]
[318]
[319]
[320]
[321]
[322]
[323]
[324]
[325]
[326]
[327]
[328]
[329]
[330]
[331]
[332]
[333]
[334]
[335]
[336]
[337]
[338]
[339]
[340]
[341]
[342]
[343]
[344]
[345]
[346]
[347]
[348]
[349]
[350]
[351]
[352]
[353]
[354]
[355]
[356]
[357]
[358]
[359]
[360]
[361]
[362]
[363]
[364]
[365]
[366]
[367]
[368]
[369]
[370]
[371]
[372]
[373]
[374]
[375]
[376]
[377]
[378]
[379]
[380]
[381]
[382]
[383]
[384]
[385]
[386]
[387]
[388]
[389]
[390]
[391]
[392]
[393]
[394]
[395]
[396]
[397]
[398]
[399]
[400]
[401]
[402]
[403]
[404]
[405]
[406]
[407]
[408]
[409]
[410]
[411]
[412]
[413]
[414]
[415]
[416]
[417]
[418]
[419]
[420]
[421]
[422]
[423]
[424]
[425]
[426]
[427]
[428]
[429]
[430]
[431]
[432]
[433]
[434]
[435]
[436]
[437]
[438]
[439]
[440]
[441]
[442]
[443]
[444]
[445]
[446]
[447]
[448]
[449]
[450]
[451]
[452]
[453]
[454]
[455]
[456]
[457]
[458]
[459]
[460]
[461]
[462]
[463]
[464]
[465]
[466]
[467]
[468]
[469]
[470]
[471]
[472]
[473]
[474]
[475]
[476]
[477]
[478]
[479]
[480]
[481]
[482]
[483]
[484]
[485]
[486]
[487]
[488]
[489]
[490]
[491]
[492]
[493]
[494]
[495]
[496]
[497]
[498]
[499]
[500]
[501]
[502]
[503]
[504]
[505]
[506]
[507]
[508]
[509]
[510]
[511]
[512]
[513]
[514]
[515]
[516]
[517]
[518]
[519]

CC THE SECOND DOUBLEBOND IN THE BIOSYNTHESIS OF 16:3 AND 18:3 FATTY
CC ACIDS, IMPROBANT CONSTITUENTS OF PLANT MEMBRANES. IT IS THOUGHT
CC TO USE PERFLUORIN AS AN ELECTRON DONOR AND TO ACT ON FATTY ACIDS
CC ESTERIFIED TO GALACTOLIPIDS, SULFOLIPIDS AND PHOSPHATIDYLGLYCEROL.
CC
CC -1- PATHWAY: POLYUNSATURATED FATTY ACID BIOSYNTHESIS
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLY)
CC -1- DEVELOPMENTAL STAGE: HIGHEST LEVELS FOUND IN EXPANDING LEAVES
CC -1- DOMAIN: THE HISTIDINE POX DOMAIN MAY CONTAIN THE ACTIVE SITE
CC AND/OR BE INVOLVED IN METAL ION BINDING.
CC
CC -1- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; U09603; AAA92800.1; -
CC EMBL; AL022199; CAI8199.1; -
CC EMBL; AL161578; CAB79813.1; -
CC EMBL; AY045621; AAK73973.1; -
CC EMBL; AY058078; AAL24186.1; -
CC EMBL; AY058452; AAL24240.1; -
CC InterPro: IPR001225; FA_desaturase.
CC Pfam: PF00487; FA_desaturase; 1.
CC ProDom: PD001081; FA_desaturase; 1.
CC
CC KW Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
CC TRANSIT peptide.
CC
CC FT CHAIN 1 69 CHLOROPLAST (BY SIMILARITY).
CC FT DOMAIN 70 448 OMEGA-6 FATTY ACID DESATURASE.
FT DOMAIN 171 175 HISTIDINE BOX-1.
FT DOMAIN 207 211 HISTIDINE BOX-2.
FT DOMAIN 367 371 HISTIDINE BOX-3.
SU SEQUENCE 448 AA, 51200 MW, 97A072CFB299F9C97 CF064,

Query Match 100.0%; Score 6; DB 1; Length 448;
Best Local Similarity 66.7%; Pred No 48;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 257 WVMWHF 262

RESULT 8

CATA_PSEPU STANDARD; PPT; 479 AA
ID CATA_PSEPU STANDARD; PPT; 479 AA
AC Q59714;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN KATA OR CATA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=303;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Corvallis;
RX MEDLINE=98019091; PubMed=9358069;
RA Kim Y.C., Miller C.D., Anderson A.L.;
RT "Identification of adjacent genes encoding the major catalase and a
RT bacterioferritin from the plant-beneficial bacterium Pseudomonas
RT putida";
RL Gene 199;219-224(1997).
CC -1- FUNCTION: FERMENTES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES
CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE
CC -1- CATALYTIC ACTIVITY: 2 H2O2(O2) = 2 H2O(O) + O2(H2O).
CC -1- COFACTOR: HEME GROUP.
CC -1- ENZYME REGULATION: ACTIVATED BY PEROXIDE

CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; U63511; AAB88219.1; -
CC HSSR; P42321; 2CAB.
CC InterPro: IPR002226; Catalase.
CC Pfam: PF00199; catalase; 1.
CC PRINTS: PR00667; CATALASE.
CC ProDom: PD000510; Catalase; 1
CC PROSITE: PS00437; CATALASE_1; 1.
CC PROSITE: PS00438; CATALASE_2; 1.
CC Oxidoreductase, Peroxidase, Iron, Heme, Hydrogen peroxide.
KW ACT_SITE 53 53 BY SIMILARITY.
FT ACT_SITE 126 126 BY SIMILARITY.
FT BINDING 336 336 PROXIMAL HEME LIGAND (BY SIMILARITY).
SU SEQUENCE 479 AA; 53381 MW; EBF3CHDE67778571 CRC64;

Query Match 100.0%; Score 6; DB 1; Length 479;
Best Local Similarity 66.7%; Pred No 50;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 209 WVMWHF 214

RESULT 9
SIK1_YEAST STANDARD; FFT; 504 AA.
ID SIK1_YEAST STANDARD; FFT; 504 AA.
AC Q12460;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2002 (Rel. 41, Last annotation update)
DE SIK1 protein (Nucleolar protein NOP56).
GN SIK1 OR NOP56 OR YIP197W OR L8167.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / YPH1;
RX MEDLINE=96040178; PubMed=7547500;
RA Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;
RT "Genetic analysis of growth inhibition by GATA-1 kappa P alpha in
RT Saccharomyces cerevisiae";
RL Cell Growth Differ. 6:789-798(1995).
RN
RN
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton D., Gattung S., Greco T., Kistner D., Kucaba T.,
RA Halleworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Mardis R., Meneses S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
RA Taich S., Trevaskis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,
RA Wohlman P., Waterston R.;
RL Submitted (SEP-1994) to the Fungal GenBank/CCP1 databases.
RN
RN
RP CHARACTERIZATION, AND MUTAGENESIS.
RX MEDLINE=98038777; PubMed=9372940;
RA Gaudier T., Reijes T., Tollervey D., Hurt F.;
RT "Nucleolar rRNA 18S-45S protein complex and its role with rRNA
RT and are required for ribosome biogenesis";
RL Mol. Cell. Biol. 17:7088-7098(1997).
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOME MATURATION AND RIBOSOME

[illegible][illegible]


```

DE 15-JUN-1999 (Ref. 15), last annotation added
DE Probable glycerol-3-phosphate acyltransferase, mitochondrial precursor
DE (EC 2.3.1.15) (GPAT).
DE F0853.2.
CC Caenorhabditis elegans.
CC Embryoid, Metazoa, Nematoda, Chordata, Subphylum, Subphylum,
CC Rhabditidae, Polychaeta, Caenorhabditis.
CC Nematoda, Nematoda.
RN 1.
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2.
RA Blandford M., Blandford H.
RE Submitted (JUL 1999) to the EMBL/GenBank/CCP databases.
CC 1- CATALYTIC ACTIVITY: Glycerol-3-phosphate + acyl-CoA ->
CC acyl-3-phosphatidylglycerol + CoA.
CC 1- PATHWAY: lipid metabolism; PHOSPHOLIPID METABOLISM; IT MAY
CC ALSO FUNCTION IN THE REGULATION OF MEMBRANE POTENTIAL.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein; Mitochondrial
CC (potential).
CC 1- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC 1- SWISS-PROT entry: P11111, 1A. Function: It is predicted to be a lipidation
CC between the Swiss Institute of Bioinformatics and the EMBL database.
CC the European Bioinformatics Institute. There are no restrictions to
CC use by non-profit institutions as long as the copyright is in the way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk.
CC
CC EMBL: U64847; AAB04876.1; -.
CC WormPep: F0853.2; CE09ABY.
DR InterPro: IPR002123; Acyltransferase.
DR Pfam: PF01553; Acyltransferase_1.
KW Phospholipid biosynthesis; transferase; Acyltransferase;
KW Transferring; with; Nucleic; Transf; [4.1.1].
FT TRANSIT 1 3 MITOCHONDRION (POTENTIAL).
FT CHAIN 1 713 PROBABLE GLYCEROL-3-PHOSPHATE
FT ACYLTRANSFERASE.
FT TRANSMEM 409 425 POTENTIAL.
SQ SEQUENCE 718 AA, 82071 MW, EDA36A4A26FC138D CPO64;
Query: WXXMHF 6
Best: 178 WXXMHF 6
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXMHF 6
Db 178 WXXMHF 183

```

```

01  "A. Erythropoietin (EPO) is a hormone that stimulates the production of red blood cells with
02  either sea anemone sodium channel toxins (isolated and amino acid
03  sequence."
04  Rb Toxin 34:57-65(1996).
05  - Function: EPOs specifically bind to the sodium channel, thereby
06  - Relating the toxin to the sodium channel, thereby
07  similarity) causes death to crabs and mice.
08  - Experimental evidence: EPOs are not secreted (by similarity).
09  - Similarity: EPOs belong to the sea anemone sodium channel inhibitory
10  TOXIN FAMILY.
11
12  CC This Swiss-Patent entry is copyright, it is produced through a collaboration
13  CC between the Swiss Institute of Bioinformatics and the EMBL Outstation.
14  CC The European Bioinformatics Institute (EBI) and the Swiss Institute of Bioinformatics (SI)
15  CC are by non-profit institutions of long standing in the field of molecular biology and
16  CC scientific information. It is not intended to be used for commercial
17  CC purposes. For more information, please contact: http://www.isb-sib.ch/announce/
18  CC or send an email to license@sib.ch.
19
20  CC EMBL: AF130344; AAF7538.1;
21  DR HSSP; P01533; LATX.
22  DR InterPro; IPR000603; Anemone_toxin.
23  DR Pfam; PF00706; Toxin4; 1.
24  DR Rb Toxin; Toxin; 34:57-65(1996) channel inhibitory, signal, cni50cyst.
25  FT SIGNAL 1 19
26  FT PROPEP 20 29
27  FT CHAIN 30 32
28  FT DISULFID 32 32 BY SIMILARITY.
29  FT DISULFID 34 34 BY SIMILARITY.
30  FT DISULFID 62 62 BY SIMILARITY.
31  FT DISULFID 80 80 BY SIMILARITY.
32  FT DISULFID 80 80 BY SIMILARITY.
33  FT DISULFID 80 80 BY SIMILARITY.
34  FT DISULFID 80 80 BY SIMILARITY.
35  FT DISULFID 80 80 BY SIMILARITY.
36  FT DISULFID 80 80 BY SIMILARITY.
37  FT DISULFID 80 80 BY SIMILARITY.
38  FT DISULFID 80 80 BY SIMILARITY.
39  FT DISULFID 80 80 BY SIMILARITY.
40  FT DISULFID 80 80 BY SIMILARITY.
41  FT DISULFID 80 80 BY SIMILARITY.
42  FT DISULFID 80 80 BY SIMILARITY.
43  FT DISULFID 80 80 BY SIMILARITY.
44  FT DISULFID 80 80 BY SIMILARITY.
45  FT DISULFID 80 80 BY SIMILARITY.
46  FT DISULFID 80 80 BY SIMILARITY.
47  FT DISULFID 80 80 BY SIMILARITY.
48  FT DISULFID 80 80 BY SIMILARITY.
49  FT DISULFID 80 80 BY SIMILARITY.
50  FT DISULFID 80 80 BY SIMILARITY.
51  FT DISULFID 80 80 BY SIMILARITY.
52  FT DISULFID 80 80 BY SIMILARITY.
53  FT DISULFID 80 80 BY SIMILARITY.
54  FT DISULFID 80 80 BY SIMILARITY.
55  FT DISULFID 80 80 BY SIMILARITY.
56  FT DISULFID 80 80 BY SIMILARITY.
57  FT DISULFID 80 80 BY SIMILARITY.
58  FT DISULFID 80 80 BY SIMILARITY.
59  FT DISULFID 80 80 BY SIMILARITY.
60  FT DISULFID 80 80 BY SIMILARITY.
61  FT DISULFID 80 80 BY SIMILARITY.
62  FT DISULFID 80 80 BY SIMILARITY.
63  FT DISULFID 80 80 BY SIMILARITY.
64  FT DISULFID 80 80 BY SIMILARITY.
65  FT DISULFID 80 80 BY SIMILARITY.
66  FT DISULFID 80 80 BY SIMILARITY.
67  FT DISULFID 80 80 BY SIMILARITY.
68  FT DISULFID 80 80 BY SIMILARITY.
69  FT DISULFID 80 80 BY SIMILARITY.
70  FT DISULFID 80 80 BY SIMILARITY.
71  FT DISULFID 80 80 BY SIMILARITY.
72  FT DISULFID 80 80 BY SIMILARITY.
73  FT DISULFID 80 80 BY SIMILARITY.
74  FT DISULFID 80 80 BY SIMILARITY.
75  FT DISULFID 80 80 BY SIMILARITY.
76  FT DISULFID 80 80 BY SIMILARITY.
77  FT DISULFID 80 80 BY SIMILARITY.
78  FT DISULFID 80 80 BY SIMILARITY.
79  FT DISULFID 80 80 BY SIMILARITY.
80  FT DISULFID 80 80 BY SIMILARITY.
81  FT DISULFID 80 80 BY SIMILARITY.
82  FT DISULFID 80 80 BY SIMILARITY.
83  FT DISULFID 80 80 BY SIMILARITY.
84  FT DISULFID 80 80 BY SIMILARITY.
85  FT DISULFID 80 80 BY SIMILARITY.
86  FT DISULFID 80 80 BY SIMILARITY.
87  FT DISULFID 80 80 BY SIMILARITY.
88  FT DISULFID 80 80 BY SIMILARITY.
89  FT DISULFID 80 80 BY SIMILARITY.
90  FT DISULFID 80 80 BY SIMILARITY.
91  FT DISULFID 80 80 BY SIMILARITY.
92  FT DISULFID 80 80 BY SIMILARITY.
93  FT DISULFID 80 80 BY SIMILARITY.
94  FT DISULFID 80 80 BY SIMILARITY.
95  FT DISULFID 80 80 BY SIMILARITY.
96  FT DISULFID 80 80 BY SIMILARITY.
97  FT DISULFID 80 80 BY SIMILARITY.
98  FT DISULFID 80 80 BY SIMILARITY.
99  FT DISULFID 80 80 BY SIMILARITY.
100 FT DISULFID 80 80 BY SIMILARITY.

```

```

Query Match      97.33%  Score 5; DB 1; Length 82;
Host Local Similarity 60.00%  Fred No 176+02;
Mismatch 0; Mismatches 0; Indels 0; Gaps 0;
      5  KXWHF 6
      1111
Db      64 KXWHF 68

```

Call for February 17, 2003, 11:04:57
 and time: 11:33:33 SRS

GenInfo version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using SW model

Run on: February 20, 2003, 11:23:27, Search time 46 seconds

(without alignment)
26.876 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 6

Sequence: 1 WXXWHF 6

Scoring table: OligoNX
Gap: 50 0, Gap: 50 0

Searched: 671540 seqs, 20604715 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671540

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: SPTREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriophage:*

17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	187	16	Q9CLS4
2	6	100.0	190	4	Q56ND5
3	6	100.0	201	10	Q9C600
4	6	100.0	202	10	Q9AXY3
5	6	100.0	213	10	Q9AXY5
6	6	100.0	216	10	Q9AXY2
7	6	100.0	219	10	Q9AK47
8	6	100.0	222	10	Q9AXY4
9	6	100.0	230	16	Q51604
10	6	100.0	241	17	Q50123
11	6	100.0	256	10	Q9PIN2
12	6	100.0	260	16	Q83250
13	6	100.0	289	17	Q8TFU3
14	6	100.0	293	16	Q41065
15	6	100.0	310	2	Q9EXF1
16	6	100.0	314	16	Q91120

17	6	100.0	381	2	Q91TQ4	Q91TQ4 agrobacteri
18	6	100.0	386	16	Q91TQ4	Q91TQ4 agrobacteri
19	6	100.0	394	10	Q9AW80	Q9AW80 guillardia
20	6	100.0	403	16	Q9YR56	Q9YR56 bruceella m
21	6	100.0	415	16	Q9YR56	Q9YR56 bruceella m
22	6	100.0	421	10	Q9C8P4	Q9C8P4 thelmutteri
23	6	100.0	423	16	Q927H4	Q927H4 listeria m
24	6	100.0	423	16	Q9Y412	Q9Y412 listeria m
25	6	100.0	424	10	Q96663	Q96663 listeria m
26	6	100.0	452	16	Q8T110	Q8T110 listeria m
27	6	100.0	452	16	Q8Z2K1	Q8Z2K1 salmonella
28	6	100.0	454	10	Q92RW0	Q92RW0 cicer ariet
29	6	100.0	462	16	Q8X5T1	Q8X5T1 escherichia
30	6	100.0	472	16	Q9YR56	Q9YR56 anabaena sp
31	6	100.0	473	5	Q9BH07	Q9BH07 leishmania
32	6	100.0	473	11	Q70106	Q70106 mus musculu
33	6	100.0	476	17	Q9HSP4	Q9HSP4 halobacteri
34	6	100.0	485	10	Q90401	Q90401 gryza gativ
35	6	100.0	497	5	Q45012	Q45012 caenorhabdi
36	6	100.0	498	16	Q9ACV6	Q9ACV6 calobacter
37	6	100.0	492	16	Q8YU21	Q8YU21 anabaena sp
38	6	100.0	494	17	Q974C6	Q974C6 sulfobobus
39	6	100.0	497	3	Q94514	Q94514 arabidopsis
40	6	100.0	499	10	Q91TV6	Q91TV6 arabidopsis
41	6	100.0	508	3	Q9P757	Q9P757 schizosacch
42	6	100.0	508	17	Q97W54	Q97W54 sulfobobus
43	6	100.0	510	5	Q9USW4	Q9USW4 drosophila
44	6	100.0	510	16	Q8YR56	Q8YR56 listeria m
45	6	100.0	511	5	Q9VW56	Q9VW56 drosophila

ALIGNMENTS

RESULT 1

Q9CLS4

ID Q9CLS4

AC Q9CLS4

DT 01-JUN-2001 (TREMBL) 17, Created

DT 01-JUN-2001 (TREMBL) 17, Last sequence update

DT 01-MAR-2002 (TREMBL) 20, Last annotation update

DE Hypothetical protein PM1135.

OS PM1135.

OC Pasteurella multocida.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

OC Pasteurella.

OX NCBI_TaxID=747;

RN [1]

FP SEQUENCE FROM N.A.

EC STRAIN EM70;

FX MEDLINE:21145866, PubMed 11248100;

PA May B.T., Zhang Q., Li L.H., Pastorian M., Whittem T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida PM70.";

PL Proc. Natl. Acad. Sci. U.S.A. 98:14450-14451(2001);

CC 1. SIMILARITY RELATIONS TO THE TRIF/ACR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

DR EMBL: AB006154; AAK03219.1; ..

DR InterPro: IPR001647; HTH_Tetr.

DR Pfam: PF00440; Tetr_1.

DR PRINTS: PR00455; HTHTR.

KW DNA-binding, Hypothetical protein, Transcription regulation;

KW Complete proteome.

SE SEQUENCE 197 AA; 23325 MW; 29E3E3E5D14115F CDS64;

Query Match: 100.0%; Score: 6, DB: 16, Length: 187;

Best Local Similarity: 66.7%; Pred. NO: 1.6e+02;

Matches: 4, Conservative: 2, Mismatches: 0, Indels: 0, Gaps: 0;

QY 1 WXXWHF 6

DB R9 WXXWHF 94

Db 154 WXXWHF 159

RESULT 6

Q9AXY2 PRELIMINARY; PRT; 216 AA.

AC Q9AXY2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Omega 6 reductase (Fragment).
GN FAD6-BN-1.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID:3708;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CV, STELLAR;
RA Fournier M., Froger N., Brunel D.; Tools designing for a genetic map
RT "Amplified consensus gene markers
RL of Arabidopsis known-function genes in Brassica";
RL Submitted (JAN-2000) to the FMBL/GenBank/DBP databases.
DR EMBL; AF229391; AAK06611;
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR ProDom; PD001081; FA_desaturase; 2.
FT NON_TER 1
FT NON_TER 216
FT NON_TER 216
SQ SEQUENCE 216 AA; 24442 MW; 1994A733D243APRA CROCA;

Query Match 100.0%; Score 6; DP 10; Length 216;
Best Local Similarity 66.7%; Pred No 1 9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
Db 157 WXXWHF 162

RESULT 7
Q94K47 PRELIMINARY; PRT; 219 AA.

AC Q94K47;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative 6-phosphogluconolactonase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CV, STELLAR;
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Raghav S., Chung M.Y.,
RA Goldsmith A.D., Lee J.M., Chuah H.L., Tang C.C., Toriumi M., Yu G.,
RA Bowers J., Carninci P., Chen H., Cheuk P., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai T., Kim C., Kosemura E.,
RA Lam P., Lin J., Meyers R.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satoh M., Seki M., Shimizu P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full length cDNA of gene K16H7.11/AT5G24400 (G110177901)";
RL Submitted (APR 2001) to the FMBL/GenBank/DBP databases
CC -1- SIMILARITY: CONTAINS 1 WD REPEAT (TRP-ASP TOWAIN)
DR EMBL; AF370305; AAK44120.1;
DR TIGRFAAS; TIGR01198; pg1; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 219 AA; 24442 MW; 1994A733D243APRA CROCA;

Query Match 100.0%; Score 6; DP 10; Length 219;

Best Local Similarity 66.7%; Pred No 1 9e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
Db 72 WARWHF 77

RESULT 8

Q9AXY4 PRELIMINARY; PRT; 222 AA.

AC Q9AXY4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Omega 6 reductase (Fragment).
GN FAD6-BO-1.
OS Brassica oleracea (cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID:3712;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CV, RAPID CYCLING;
RA Fournier M., Froger N., Brunel D.;
RT "Amplified consensus gene markers; Tools designing for a genetic map
RL of Arabidopsis known-function genes in Brassica";
RL Submitted (JAN 2000) to the FMBL/GenBank/DBP databases.
DR EMBL; AF229391; AAK06611;
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR ProDom; PD001081; FA_desaturase; 2.
FT NON_TER 1
FT NON_TER 222
FT NON_TER 222
SQ SEQUENCE 222 AA; 25761 MW; A612F53J28U24E CROCA;

Query Match 100.0%; Score 6; DP 10; Length 222;
Best Local Similarity 66.7%; Pred No 1 9e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
Db 163 WXXWHF 168

RESULT 9
O53604 PRELIMINARY; PRT; 230 AA.

AC O53604;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein RV0059.
GN RV0059 OR MT0030 Q2 OR MT0065.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycetaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1773;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE:9825987; PubMed:9634230;
RA Cole S.T., Brosch P., Parkhill J., Garfield T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock P., Bashir T., Brown T., Chillingworth T., Collins F.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Maitland S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Whittall B.G.;
PT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";

[illegible]

Query Match	100.0	Score 0	bp 16	Length 300
Best Local Similarity	66.7%	Prod. No. 1.92102		
Matches	4	Conservation	0	Weighted

Q7.	1	EXCERPT	5
Q8.	26	WITNESS	21

RESULT 0
059123
+D 059123 PRELIMINARY; PRT; 241 AA.

DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
DT 01-MAR-2002 (TREMblrel. 09, last annotation update)
DE Hypothetical protein PH1454.
GN PH1454.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI Taxid=53953;
RN (1)
RP SEQUENCE FROM N.A.
RC SPAIN-OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawai, Yuji Y., Kawada N., Horiuchi H., Horiuchi Y., Horiuchi T.,
RA Tanabe, S., Sekine M., Eki S. I., Kasegi H., Hosojima S., Nagai T.,
RA Imai M., Ohta M., Iizawa K., Nakai H., Nakamura M., Nakamura M.,
RA Funahashi T., Tanaka T., Kishi Y., Kamekura S., Kamekura T., Oguchi A.,
RA Aoki K. I., Yamanaka T., Nakamura Y., Shibata T., Horiuchi H.,
RA Matsuda H., Shiga H., Horiuchi H.,
RA Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeon, Pyrococcus islandicus.
RI DNA Base: 358-571899.
RI CNA Base: 358-571899.
DR EMBL; AF030006; BAB30561.1; F00000
PC Pyrococcus horikoshii strain Dsm 9790
SQ SEQUENCE 241 AA; 26494 MW; A095810EAE1E70818 CD0304;

Every Match	100.0%	Score 67	(h 17)	Length 241
Best Local	98.0%	Epoch 100	+ 30000	
Modes	4	Conservative	2	Mematches 0
				Labels 0
				Caps 0

QY	1	WXXWHF	6
DB	193	WQTWHF	198

RECORD 11		
09FIN2		
ID 09FIN2	PREDICTION	FRT, 236 AM.
AC 09FIN2		
DT 01-MAR-2001	(TIMBLR1, 16, Created)	
DT 01-MAR-2001	(TIMBLR1, 16, Last sequence update)	

[illegible]

SEQUENCE FROM N. A.

PY: MEDLINE-00156222; PubMed-10048438;

RA Tabata S.;

[illegible]

Physiology assigned P1 and TAC clones. ");

DR EMEL; AB016884; BAR11233.1; -.

1. Fluoridation of drinking water.

[illegible]

0
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

	Query Match	100.0%	Score 6	DE 10	Length 256
Best Local Similarity	66.7%		Pred. No. 2e+02		
Matched	4	Consistent	0	Mismatched	0

QY	1	WXXWHF	6
		: :	
Db	72	WARWHF	77

RESULT 12
083250
10 000000
FROM MOUNT;
END; 000000

```

DT 01-NOV-1998 (TREMblrel. 08, Created)
BT 01-NOV-1998 (TREMblrel. 08, last sequence update)
BT 01-FEB-1992 (TREMblrel. 01, last annotation update)
DE D, D-carboxypeptidase, putative.
GN TP0221.
OS Treponema pallidum.
OC Bacteria, Spirochaetales, Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
R1 SEQUENCE FROM N.A.
R2 STRAIN=NICHOLS;
R3 WELSH, J. M., HARRIS, R. T., WELSH, J. M., WHITE, C., ELLIS, G. G.,
R4 FRANCIS, E., GILLES, W., HICKY, E. K., OLIVER, R., KOSMAN, N. A.,
R5 FRANCIS, E., GILLES, W., HICKY, E. K., OLIVER, R., KOSMAN, N. A.,
R6 Schlegel, E., Haddad, M., Wilson, M. D., Calzberg S., Peterson T.,
R7 Malak, H., Richardson, P., Howell, K., Ollabar, M., Wrethback T.,
R8 McDonald, M., Arnold, P., Egan, C., Cotton, M. D., Fujii, C., Garland S.,
R9 Hatch, E., Host, K., Roberts K., Sankaly, M., Weidman U., Smith H. O.,
R10 Venter, J. C.)
R11 "complete genome sequence of Treponema pallidum, the syphilis
R12 spirochete."
R13 Science 281:375-388(1998) .
R14 EMBL: A8001704; AAC65208.1; -
R15 FICR, F0221; -
R16 InterPro; IPR000561; EGF-like.
R17 InterPro; IPR003709; Vany.
R18 Pfam; PF02557; Vany; 1.
R19 Prosite; PS00022; EGF_1; UNKNOWN_1.
R20 Carboxypeptidase; complete proteome.
R21 SEQUENCE 280 AA; 2933 MW; 1E96014E122FF6 CRC64;
SV

```

```

Query Match:      100.0%; Score 6; DB 16; Length 260;
Best local Similarity 66.7%; Pred. No. 2.1e+02;
Matches      4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY          1 WXXWHF 5
|:|:|

```


Db 220 WPIWHF 225

RESULT 13

Q8TRU3

AC 08TRU3; PRELIMINARY; PRT; 299 AA

DT 01-JUN-2002 (TREMURel 21, Created)

DT 01-JUN-2002 (TREMURel 21, Last sequence update)

DT 01-JUN-2002 (TREMURel 21, Last annotation update)

DE Hypothetical protein MA1162.

GN MA1162.

OS Methanosarcina acetivorans.

OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;

OC Methanosarcinaceae; Methanosarcina.

OX NCBI_TaxID=2214;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN C3A / ATCC 35935 / DSM 2934.

RX MEDLINE 21929760; PubMed 11932338;

RA Galagan J E., Musbaum C., Roy A., Endrizzi M G., Macdonald P.,

RA Fitzhugh W., Galin S., Engle P., Smirnov S., Atwood D., Brown A.,

RA Allen N., Nayler J., Ganga-Thamm H., DeArnell E., Johnson P.,

RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,

RA Zimmer A., Barber R D., Cann I., Graham D E., Grahame D A., Guss A M.,

RA Hedderich P., Ingram-Smith C., Kuettner H C., Krzycki J A.,

PA Leigh J A., Li W., Liu T., Muthofsky R., Peave J.N., Smith K.,

PA Springer T A., Teyam I A., White O., White P H., de Maessene E C.,

RA Ferry J.G., Jarell K F., Ding H., Macario A D L., Paulsen I.,

RA Pritchett M., Sowers K.P., Swanson R.V., Zinder S.H., Lander E.,

RA Metcalf W.W., Birren B.;

RT "The genome of Methanosarcina acetivorans reveals extensive metabolic

RT and physiological diversity";

RL Genome Res 12:432-442(2002)

DR EMBL; AB010783; AAM04583.1;

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 299 AA; 33299 MW; E422100BFA52B2E6 CRC64;

Query Match 100.0%; Score 6; DB 17; Length 289;

Best local similarity 66.7%; Pred. No. 2.3e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

Db 168 WSIWHF 173

RESULT 14

Q91065

ID Q91065; PRELIMINARY; PRT; 293 AA.

AC Q91065;

DT 01-MAR-2001 (TREMURel 16, Created)

DT 01-MAR-2001 (TREMURel 16, Last sequence update)

DT 01-JUN-2002 (TREMURel 21, Last annotation update)

DE Hypothetical protein PA2778.

GN PA2778.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.T., Britman F.S.B., Hutnagle W.O., Kowalik D.T., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.P., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock P.F.W., Iori S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

DR EMBL; AF004705; AAG06166.1;

DR InterPro; IPP005074; Peptidase_C39

DR InterPro; IPR001440; TPR.

DR Pfam; PF03412; Peptidase_C39; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 293 AA; 32916 MW; 0E4410BEEBFA0DE1 CRC64;

Query Match 100.0%; Score 6; DB 16; Length 293;

Best local similarity 66.7%; Pred. No. 2.3e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

Db 101 WPIWHF 106

RESULT 15

Q9EXF1

AC Q9EXF1; PRELIMINARY; FFT; 310 AA.

DT 01-MAR-2001 (TREMURel 16, Created)

DT 01-MAR-2001 (TREMURel 16, Last sequence update)

DT 01-JUN-2002 (TREMURel 21, Last annotation update)

DE Putative phosphotransferase D-arabitol specific component 110

DE (Fragment).

GN ATLC.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacilli; Clostridium group; Bacillales;

OC Listeriaceae; Listeria.

OX NCBI_TaxID=1639;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LO29;

RX MEDLINE=2122973; PubMed 11316371;

RA Saklani-Jusforgues H., Fontan E., Goossens P.L.;

RT "Characterisation of a Listeria monocytogenes mutant deficient in D-

RT arabitol fermentation.";

RL Res. Microbiol. 152:175-177(2001).

DR EMBL; AJ292552; CAC20641.1;

DR InterPro; IPR004703; Gal_spec_11C.

DE Pfam; PF03611; E11C_GAT_1.

KW Transferase.

FT NON_TER 1 1

FT NON_TER 310 310

SQ SEQUENCE 310 AA; 33711 MW; A2F9FE9C0C8DA0F CRC64;

Query Match 100.0%; Score 6; DB 2; Length 310;

Best local similarity 66.7%; Pred. No. 4.4e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6

Db 18 WNIWHF 23

Search completed: February 20, 2003, 11:28:57

Job time: 46 secs



GenCore version 5.1.3
Copyright 1993-2003 CompuGen Inc.

OM protein - protein search, using sw model

Run on: February 20, 2003, 11:21:07, Search time 29.333 seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-11

Perfect score: 6
Sequence: 1 WXXWHP 6

Scoring table: OLIGOEX
Gapop 60 0, Gapext 60 0

Searched: 908470 seqs, 13320620 residues

Word size: 0

Total number of hits satisfying chosen parameters: 508470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post processing: listing first 45 summaries

Database:

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	100.0	6	21 AAB01492	Peptide which bind
2	6	100.0	6	21 AAB01499	Peptide which bind
3	6	100.0	6	21 AAB01505	Peptide which bind
4	6	100.0	6	21 AAB01506	Peptide which bind
5	6	100.0	36	22 AAB01893	Peptide #8399 ento
6	6	100.0	36	22 AAM61753	Human brain expres
7	6	100.0	36	22 AAM74549	Human bone marrow
8	6	100.0	36	22 AAM74663	Peptide #8700 ento
9	6	100.0	43	22 AAT14693	Novel bone marrow
10	6	100.0	43	23 ABG64892	Human albumin fusi

11	6	100.0	53	23 AAE21233	Human gene 18 ento
12	6	100.0	57	23 AAM61758	Protein:protein
13	6	100.0	65	24 ABE12322	Human GEF 1, cell cycle
14	6	100.0	87	22 AAU56391	Protonibacterium
15	6	100.0	96	22 AAU14777	Novel bone marrow
16	6	100.0	119	23 APP06170	Human GEF protein
17	6	100.0	173	22 AAC60684	lipid modification
18	6	100.0	173	22 AAC68334	Human eucalyptus
19	6	100.0	175	22 ABB03211	Human polyprotein
20	6	100.0	175	22 AAM42371	Human polyprotein
21	6	100.0	267	21 AAG09463	Arabidopsis thalia
22	6	100.0	266	21 AAG44075	Arabidopsis thalia
23	6	100.0	266	21 AAG48293	Arabidopsis thalia
24	6	100.0	269	22 AAG02313	Arabidopsis thalia
25	6	100.0	306	21 AAG05810	Arabidopsis thalia
26	6	100.0	306	21 AAG47194	Arabidopsis thalia
27	6	100.0	324	21 AAG05215	Arabidopsis thalia
28	6	100.0	325	21 AAG09467	Arabidopsis thalia
29	6	100.0	325	21 AAG44074	Arabidopsis thalia
30	6	100.0	325	21 AAG48292	Arabidopsis thalia
31	6	100.0	338	21 AAG05809	Arabidopsis thalia
32	6	100.0	338	21 AAG47193	Arabidopsis thalia
33	6	100.0	376	19 AAM46287	Rhizobium meliloti
34	6	100.0	377	19 AAM46288	Brucella abortus D
35	6	100.0	418	15 AAB60501	linoleic acid desa
36	6	100.0	423	21 AAG09465	Arabidopsis thalia
37	6	100.0	423	23 AAB49593	listeria monocytog
38	6	100.0	424	22 AAE13431	coybean microsomal
39	6	100.0	427	22 AAU20451	Human secreted pro
40	6	100.0	427	22 AAG01833	Novel human scpla
41	6	100.0	448	21 AAG05808	Arabidopsis thalia
42	6	100.0	448	21 AAG47192	Arabidopsis thalia
43	6	100.0	448	22 ABR02179	herbicide 1177 arvic
44	6	100.0	452	21 AAG44623	Arabidopsis thalia
45	6	100.0	466	22 AAB05611	Human Protein regu

ALIGNMENTS

RESULT 1
ID AAB01492 standard; peptide; 6 AA.
AC AAB01492;
XX
XX 09-NOV 2000 (first entry)
XX
XX Peptide which binds to transcription factor EGF 1 DNA binding domain.
XX
XX DNA binding; transcription factor, EGF, EGF 1, cell cycle; DP-1;
XX activation; transcription; apoptosis; proliferative disorder;
XX psoriasis; restenosis.
XX
XX Synthetic.
XX
XX WO200044771-A1.
XX
XX 03-AUG-2000.
XX
XX 26 JAN 2000; 2000WO-0900227
XX
XX 26 JAN 1999; 99GB-0001710.
XX
XX (PROL-) PROLIFIX LTD.
XX
XX Mueller R, Kontermann RE, Montigiani S;
XX
XX WPI; 2000 512806/48.
XX
XX Peptides binding to the DNA binding domain of transcription factor EGF
XX and inhibiting cell cycle progression; useful for the treatment of
XX cancer

[illegible][illegible]

DE Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
XX DNA binding; transcription factor, E2F, E2F-1, cell cycle, DP-1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN WO200044771-A1.
XX
XX 03-AUG-2000.
PD
XX
XX 26-JAN-2000; 2000US-060227.
PF
XX
XX 26-JAN-1999; 99GB-0001710.
PR
XX
XX (PROL-) PROLIFIX LTD.
PA
XX
XX Mueller R, Kontermann RE, Montigiani S;
PI
XX
XX WPI; 2000-532806/48
DR
XX
XX Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
XX Example; Page 26; 42pp; English.
PS
XX
XX Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 6; DB 21; Length 6;
Best Local Similarity 66.7%; Pred. No. 7.8e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 1 WVAWHF 6

RESULT 5
ABR40893
ID ABR40893 standard; Peptide; 36 AA.
XX
XX ABR40893;
AC
XX
XX 04-FEB-2002 (first entry)
DT
XX
XX Peptide #8399 encoded by human foetal liver single exon probe.
DE
XX
XX Human; foetal liver; gene expression, single exon nuclear acid probe.
KW
XX
XX Homo sapiens
OS
XX
XX WO200157277-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX
XX 26-MAY-2000; 2000US-0207456.
PR
XX
XX 30-JUN-2000; 2000US-0608408.
PR
XX
XX 03-AUG-2000; 2000US-0632366.
PR
XX
XX 21 SEP 2000; 2000US-0234687.
PR
XX
XX 22 SEP 2000; 2000US-0236359.
PR
XX

PP 04-OCT-2000; 2000GB-0004263.
XX
XX (MOLE) MOLECULAR DYNAMICS INC.
FA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-483447/52.
DR
XX
XX Human genome derived single exon nuclear acid probes useful for
PT analyzing gene expression in human fetal liver
PT
XX
XX Claim 27, SEQ ID NO 33528, 639bp + sequence listing, English.
PS
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp:wipo.int/pub/published_pat_sequences.
XX
XX
SQ Sequence 36 AA;

Query Match 100.0%; Score 6; DB 22; Length 36;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHF 6
|:||||
Db 19 WRAWHF 24

RESULT 6
AAM61753
ID AAM61753 standard; Protein; 36 AA.
XX
XX AAM61753;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 33858.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
XX
XX Homo sapiens.
OS
XX
XX WO200157275 A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX
XX 26-MAY-2000; 2000US-0207456.
PR
XX
XX 30-JUN-2000; 2000US-0608408.
PR
XX
XX 03-AUG-2000; 2000US-0632366.
PR
XX
XX 21 SEP-2000; 2000US-0234687.
PR
XX
XX 27-SEP-2000; 2000US-0236359.
PR
XX
XX 04-OCT-2000; 2000GB-0024263.
PA
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001 483446/52.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains
PT
XX

PS Example 4: SEQ ID NO: 33858; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleotide acid

CC probes which are derived from genetic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is a protein encoded by one of

CC the probes of the invention.

XX

SQ Sequence 36 AA;

Query Match 100.0%; Score 6; DB 22; Length 36;

Best Local Similarity 66.7%; Prod. No. 25;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHP 6

DB 19 WRAWHP 24

RESULT 7

AAM74549

ID AAM74549 standard; Protein; 36 AA.

XX

AC AAM4549;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 4549.

XX

EN Human bone marrow expressed probe encoded protein SEQ ID NO: 4549.

XX

OS Homo sapiens.

XX

PN M0200157272 A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001MO-US00668.

XX

PR 04-FEB-2000; 2000MO-0180212.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0207456.

PR 01-AUG-2000; 2000US-0512366.

PR 21-SEP-2000; 2000US-0236359.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (NOTE) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488907/53.

XX

PT Human genome derived single exon nucleotide acid probes useful for

PT analyzing gene expression in human bone marrow

XX

PS Example 4; SEQ ID NO: 34855; 650bp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleotide acid

CC probes which are derived from genetic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is a

CC protein encoded by one of the probes of the invention.

XX

SQ Sequence 36 AA;

Query Match 100.0%; Score 6; DB 22; Length 36;

Best Local Similarity 66.7%; Prod. No. 25;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHP 6

DB 19 WRAWHP 24

RESULT 8

AAM34663

ID AAM34663 standard; Protein; 36 AA.

XX

AC AAM34663;

XX

DT 17-OCT-2001 (first entry)

XX

DE Expressed in placenta by probe for measuring placental gene expression.

XX

EN Expressed in placenta by probe for measuring placental gene expression.

XX

OS Homo sapiens.

XX

PN M0200157272-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001MO-US00663.

XX

PR 04-FEB-2000; 2000US-0180212.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0207456.

PR 01-AUG-2000; 2000US-0512366.

PR 21-SEP-2000; 2000US-0236359.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (NOTE) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488907/53.

XX

PT Human genome derived single exon nucleotide acid probes useful for

PT analyzing gene expression in human placenta

XX

PS Example 4; SEQ ID NO: 34855; 650bp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleotide acid probes (SHNP)

CC for analyzing gene expression in placenta. The present sequence is a protein encoded by one

CC of such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of

CC human genetic disorders.

XX

SQ Sequence 36 AA;

Query Match 100.0%; Score 6; DB 22; Length 36;

Best Local Similarity 66.7%; Prod. No. 25;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWHP 6

DB 19 WRAWHP 24

RESULT 9

AAM14693

ID AAM14693 standard; Protein; 43 AA

XX

AC AAM14693;

XX

DT 24-OCT-2001 (first entry)

XX

DE Novel bone marrow polypeptide #92.

XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX Homo sapiens.
OS
XX W0200157187-A2
PN
XX
PD 09-AUG-2001
XX
PF 05-FEB-2001; 2001WO-US03782
XX
PR 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 30-NOV-2000; 2000US-0660683.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Fort 1F, Poylo RJ, Tang YT, Liu C, Acundi V, Zhou F, Xue AJ,
PI Ren F, Drmanac RT;
XX
DR WPI: 2001-488875/53.
DR N PQR: AAS22988
XX
PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX
PS Claim 10; Page 248; 392pp; English.
XX
CC AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as RNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX
SQ Sequence 43 AA,
OY
Query Match 100.0%; Score 6; DB 22; Length 43;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 14 WLIWHF 19
OY 1 WXXWHF 6
Db 14 WLIWHF 19
RESULT 10
ABG64892
ID ABG64892 standard; Protein; 53 AA.
XX
AC ABG64892;

XX
DT 27-AUG-2002 (first entry)
XX
DE Human albumin fusion protein #1567.
XX
KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW human serum albumin; HSA; cancer; reproductive disorder;
KW digestive disorder; immune disorder; endocrine disorder;
KW haematopoietic disorder; neural disorder; connective disorder;
KW cytostatic; antiinfectivity; antiinflammatory; antitumor;
KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;
KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW osteopathic; antiarthritic.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX W0200177137-A1.
XX
PN 18-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US11988.
XX
PR 12-APR-2000; 2000US-229358P.
PR 25-APR-2000; 2000US-199384P.
PR 21-DEC-2000; 2000US-256931P.
XX
PA (HUMA) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Haseltine WA;
PI
XX
DR WPI: 2002-010886/01
XX
PT New fusion protein for treating disease e.g. diabetes comprising an
PT albumin fused to a therapeutic protein -
XX
PS Claim 1; Page 1582; 2102pp; English.
XX
CC The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA, also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or
CC disorder that may be modulated by therapeutic protein X. The albumin
CC extends the shelf life of protein X, and may increase its biological
CC in vitro/in vivo activity. The protein is useful for treating and
CC diagnosing disorders such as cancer, reproductive disorders, digestive
CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
CC (e.g. diabetes), haematopoietic disorders, neural disorders
CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
CC fusion proteins of the invention.
XX
SQ Sequence 53 AA;
OY
Query Match 100.0%; Score 6; DB 23; Length 53;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 21 WWSWHF 26
OY 1 WXXWHF 6
Db 21 WWSWHF 26
RESULT 11
AAE21233
ID AAE21233 standard; Protein; 53 AA.
XX
AC AAE21233;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human gene 18 encoded secreted protein HNNBM45, SEQ ID NO:98.

XX		Human; secreted protein; immune disorder; anti-allergic; antirheumatic;
KW		rheumatoid arthritis; breast neoplasia; breast cancer; antiarthritic;
KM		neurological disease; Alzheimer's disease; Parkinson's disease; trauma;
KM		multiple sclerosis; encephalitis; cysticercosis; leishmaniasis; malaria;
KM		anti-inflammation; phthalocyanine; dermatological; immunomodulatory;
KW		immune activity; immunosuppressive; antitubercular; infectious;
KW		gene therapy; autoimmune disease; Huntington's disease; peritonitis;
KW		demyelinating disease; peripheral neuropathy; congenital malformation;
KW		spinal cord injury; peripheral neuropathy; ischaemia; perception;
KW		multiple sclerosis; infection; haemorrhage; schizophrenia; dementia;
KW		depression; panic disorder; learning disability; AIDS; feeding disorder;
KW		hypertrophic pyloric stenosis; sleep pattern; cardiovascular disorder;
KW		reproductive disorder; digestive system disorder; behavioural disorder.
XX		
OS	Homo sapiens.	
FH	Key:	location/Qualifiers
FT	Peptide	1..31
FT		/label= Signal_peptide
FT	Protein	32..63
FT		/label= Mature_secreted_protein
FT	Misc. Difference	02
FT		/label= Unknown
FT		/note= "Encoded by NCA"
PX	WO200216390-A1.	
XX		
PD	28-FEB-2002.	
XX		
PF	17-JAN-2001; 2001WO-US01435.	
XX		
FE	13-MAR-2000; 1000US 126387P.	
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Farnham-Storke M, Baker AJ, Blase DE, Lippert JB, Liang HD,	
PT	Moore PA, Wolf T, Ehrner P, Dean DP, Shi Y, Choi CH, Fischella M,	
PI	Ni J.	
XX		
DR	WP1; 2002-304113/34.	
DR	N-PDB; AAD33734.	
XX		
PT	An isolated polypeptide (A) comprising a polynucleotide which	
PT	encodes a polypeptide useful in the diagnosis and treatment of	
PT	disorders e.g. immune disorders -	
XX		
PS	Claire H, Page DC, Duggan, English.	
XX		
CC	AAD33692-AAD33736 represent cDNAs corresponding to 21 human secreted	
CC	protein genes, and AAE21191-AAE21235 represent the proteins they encode.	
CC	AAE21191-AAE21235 represent human secreted protein fragments. The genes	
CC	and their corresponding secreted proteins are useful for preventing,	
CC	treating or curing various diseases and disorders, such as immune system	
CC	therapy. EMBL JI accession numbers for the sequences encoding the	
CC	secreted proteins are listed in Table I. By determining the presence of	
CC	a specific gene in the genome, specific diseases can be identified in one of the	
CC	21 genes, based on the sequence of each gene, they are also highly relevant,	
CC	and include potential targets for the design of a therapeutic agent.	
CC	Immune or autoimmune diseases e.g. AIDS, acquired immune deficiency	
CC	syndrome, asthma, sarcoid and rheumatoid arthritis, breast neoplasia	
CC	and breast cancer, neurological diseases e.g. Alzheimer's disease,	
CC	Parkinson's disease, Huntington's disease, multiple sclerosis,	
CC	amyotrophic lateral sclerosis, peripheral neuropathy, perianitis,	
CC	trauma, congenital malformations, spinal cord injuries, etc.	
CC	Neurofibrils induced by neurotoxins, peripheral neuropathies, multiple	
CC	sclerosis, ischaemic and infarction, haemorrhages, stroke, thrombosis,	
CC	dementia, hepatitis, liver disorders, learning disabilities, AIDS,	
CC	altered behaviours e.g. disorders in feeding, sleep patterns, behaviour	
CC	and perception, emphysema, etc., disorders in heart rate, blood pressure,	
CC	reproduction and reproductive system, behaviour and its effects and	
CC	infection, inflammation, etc. The present sequence represents a human	
CC	secreted protein of the invention.	

[illegible]

XX Sequence 57 AA;
SQ Query Match 100.0%; Score 6; DB 23; Length 57;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXXWHF 6
Db 4 WFPWHF 9
RESULT 13
ABP07322
ID ABP07322 standard; Protein; 65 AA.
XX
AC ABP07322;
XX
DT 24 JUN 2002 (first entry)
XX
DE Human CRFX protein sequence SEQ ID NO:14626.
XX
KW Human; open reading frame; CRFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis
XX
OS Homo sapiens
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 23-AUG-2000; 2000US-228718P.
XX
PA (CRFX) CRPAGEN CORP
XX
PI Shimkova RA, Leach MD;
XX
DR WPI; 2002 106308/14
DR N-PSDB; ABN23074.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosis,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders
XX
PS Disclosure; SEQ ID 14626; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, CRFX, where X is 1-11491 (see Table 1
CC in the specification). APP1500 to APP1500 encode the human CRFX
CC proteins given in ABP00010 to APP1500. CRFX proteins are useful for
CC treating or preventing a pathology associated with an CRFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with CRFX-associated disorder. CRFX polynucleotide
CC sequences can be used in gene therapy. CRFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. CRFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WPI
CC at ftp://wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 65 AA;
QY 1 WXXWHF 6
Db 27 WKSWHF 32
Query Match 100.0%; Score 6; DB 23; Length 65;
Best Local Similarity 66.7%; Pred. No. 41;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
RESULT 14
AAU56391
ID AAU56391 standard; Protein; 87 AA.
XX
AC AAU56391,
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #17287.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; FUSIA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathy; neuroprotection.
XX
OS Propionibacterium acnes.
XX
FN WO200181541-A2
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US 208841P.
PR 07-JUL-2000; 2000US 216747P.
XX
FA (CCRI) CCPIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'Watsoneuve S, Chang Y, Tan C, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59575.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID NO 17586; 1069pp; English.
XX
CC Sequences AAU39105-AAU64017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC	down-regulate expression and activity of P. across polypeptides and
CC	therefore, treat diseases infections. The antibodies may also be used to
CC	diagnose diseases for determining disease presence, for example, by
CC	enzyme-linked immunosorbent assay (ELISA).
CC	Note: The sequence ID# for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at http://wipo.int/pat/publicated_pat_sequences .
XX	Sequence 97 AA.
QY	Query Match:
DB	Best Local Similarity 66.78% E-Value 1E-22 Length 61,
	MATCHES 4, Conservation 2, Mismatches 0, Indels 0, Gaps 0,
	WPMWF 32
RESULT 15	
AU14777	
ID AU14777 standard; Protein; 96 AA.	
AC AU14777;	
DT 24-OCT-2001 (first entry)	
DE Novel bone marrow polypeptide #176.	
KM Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;	
KM hematopoiesis; myeloid; lymph cell disorder; tissue regeneration;	
KM wound healing; nutritional supplement; immune disorder;	
KM severe combined immunodeficiency; SCID.	
OS Homo sapiens.	
PN MO200157187-A2.	
PD 09-AUG-2001.	
PE 05-FEB-2001; 2001WO/0503782.	
PR 03-FEB-2000; 2000US-0496914.	
PR 20-JUN-2000; 2000US-0598075.	
PR 19-JUN-2000; 2000US-0620325.	
PR 30-NOV-2000; 2000US-0250683.	
XX (HYSE-) HYSEQ INC.	
PI Fold CF, Doyle BJ, Tang YT, Liu C, Roudi V, Zhou F, Xue AJ;	
PI Ren F, Dymnac RT;	
DR WPI; 2001-488975/53.	
DR N FSDP; AAS23082.	
PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic	
PT and gene therapy -	
PS Claim 10; Page 137; 192pp; English.	
XX AU14777 represent novel bone marrow polypeptides of the	
XX invention. The predicting and corresponding coding sequences may be used	
CC in the prevention, diagnosis and treatment of diseases associated with	
CC inappropriate bone marrow polypeptide expression. For example, to treat	
CC disorders associated with decreased expression by rectifying mutations	
CC or deletions in a patient's genome that affect the activity of the	
CC polypeptides by expressing inactive proteins to supplement the	
CC patient's own production of the polypeptide. Additionally, the nucleic	
CC acids may be used to produce the polypeptides, by inserting the nucleic	
CC acids into a host cell and culturing the cell to express the protein.	
CC The nucleic acid and its complementary sequences may also be used as DNA	
CC probes in diagnostic assays to detect and quantitate the presence of	
CC similar nucleic acid sequences in samples, and therefore which patients	

01	may be in good preventative therapy. The proteins may also be used as
02	antigens in the prophylactic and therapeutic against the cancer proteins
03	and in assays. Identifying kind is of their expression and activity.
04	The anti bodies narrow protein antibodies and antigens may also be used
05	to down regulate expression and activity. The antibodies may also be used
06	as diagnostic agents for detecting the presence of the protein in samples
07	such as by enzyme linked immunosorbent assay (ELISA). The proteins
08	may be used to regulate hematopoietic activity, and consequently in the
09	treatment of myeloid or lymph cell disorders, in tissue regeneration,
10	such as wound healing, as a nutritional supplement, and in treatment of
11	immune disorders such as severe combined immunodeficiency (SCID).
XX	
XX	Sequence 96 AA;
01	Query Match
02	Best Local Similarity 66.7%; Pred. No: 57;
03	Matches 4, Conservative 2, Mismatches 0, Totals 0, Gaps 0,
QY	1 WXXWHF 6
DB	67 WLIWHF 72

Search completed: February 20, 2003, 11:26:32
Job time : 29.3333 secs

GenCore version 5.1.3
Copyright 1993 - 2000 Genengen Ltd.

OM Protein - Protein search using SW model

File: us-09-912-414-9 February 20, 2003, 15:25:48, Search time: 24.333 seconds
(without alignments)
27.256 Million cell updates/sec

Title: US-09-912-414-9

Perfect score: 31

Sequence: 1 WXXWXP 6

Scoring table:

BLOSUM62X Gapop 10.0, Gapext 0.5

Searched: 900470 seqs, 13350620 residues

Total number of hits satisfying chosen parameters: 308470

Minimum DB seq length: 0

Maximum DB seq length: 20000000

Post processing: Minimum Match: 94

Maximum Match: 100%

Listing first 45 summaries

Database: A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneq-emb1/AA1980.DAT*
2:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1981.DAT*
3:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1982.DAT*
4:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1983.DAT*
5:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1984.DAT*
6:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1985.DAT*
7:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1986.DAT*
8:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1987.DAT*
9:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1988.DAT*
10:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1989.DAT*
11:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1990.DAT*
12:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1991.DAT*
13:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1992.DAT*
14:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1993.DAT*
15:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1994.DAT*
16:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1995.DAT*
17:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1996.DAT*
18:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1997.DAT*
19:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1998.DAT*
20:	/SIDS2/gcgdata/geneq/geneq-emb1/AA1999.DAT*
21:	/SIDS2/gcgdata/geneq/geneq-emb1/AA2000.DAT*
22:	/SIDS2/gcgdata/geneq/geneq-emb1/AA2001.DAT*
23:	/SIDS2/gcgdata/geneq/geneq-emb1/AA2002.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	6	AA57386	Peptide for treati
2	31	100.0	6	AA57386	Peptide for treati
3	31	100.0	6	AA57386	Peptide for treati
4	31	100.0	6	AA57386	Peptide for treati
5	31	100.0	6	AA57386	Peptide for treati
6	31	100.0	6	AA57386	Peptide for treati
7	31	100.0	6	AA57386	Peptide for treati
8	31	100.0	6	AA57386	Peptide for treati
9	31	100.0	6	AA57386	Peptide for treati
10	31	100.0	6	AA57386	Peptide for treati

11	31	100.0	8	15	AA57386	Peptide for treati
12	31	100.0	9	21	AA57386	Peptide for treati
13	31	100.0	10	15	AA57386	Peptide for treati
14	31	100.0	10	15	AA57386	Peptide for treati
15	31	100.0	11	19	AA57386	Peptide for treati
16	31	100.0	11	21	AA57386	Peptide for treati
17	31	100.0	12	15	AA57386	Peptide for treati
18	31	100.0	12	21	AA57386	Peptide for treati
19	31	100.0	12	21	AA57386	Peptide for treati
20	31	100.0	12	18	AA57386	Peptide for treati
21	31	100.0	13	23	AA57386	Peptide for treati
22	31	100.0	14	22	AA57386	Peptide for treati
23	31	100.0	15	21	AA57386	Peptide for treati
24	31	100.0	15	21	AA57386	Peptide for treati
25	31	100.0	15	23	AA57386	Peptide for treati
26	31	100.0	15	23	AA57386	Peptide for treati
27	31	100.0	15	23	AA57386	Peptide for treati
28	31	100.0	17	21	AA57386	Peptide for treati
29	31	100.0	19	22	AA57386	Peptide for treati
30	31	100.0	20	23	AA57386	Peptide for treati
31	31	100.0	20	23	AA57386	Peptide for treati
32	31	100.0	20	23	AA57386	Peptide for treati
33	31	100.0	21	22	AA57386	Peptide for treati
34	31	100.0	21	22	AA57386	Peptide for treati
35	31	100.0	22	22	AA57386	Peptide for treati
36	31	100.0	22	22	AA57386	Peptide for treati
37	31	100.0	22	22	AA57386	Peptide for treati
38	31	100.0	22	22	AA57386	Peptide for treati
39	31	100.0	22	22	AA57386	Peptide for treati
40	31	100.0	22	23	AA57386	Peptide for treati
41	31	100.0	22	23	AA57386	Peptide for treati
42	31	100.0	22	23	AA57386	Peptide for treati
43	31	100.0	22	23	AA57386	Peptide for treati
44	31	100.0	22	23	AA57386	Peptide for treati
45	31	100.0	22	23	AA57386	Peptide for treati

ALIGNMENTS

RESULT 1
AA57386
ID AA57386 standard; peptide; 6 AA.
AC AA57386;
XX 21-MAP-1995 (first entry)
XX Peptide for treating diseases related to anti-DNA antibodies.
XX Carrier, absorbing agent, treatment, anti-DNA antibodies, for anti-DNA.
XX Synthetic.
XX JP06192290-A.
XX 12-JUN-1994.
XX 18-JAN-1993; 93JP-0006098.
XX 30 SEP 1992, 92JP 0261821.
XX (KURS) KURARAY CO LTD.
XX WPI, 1994 260519/22.
XX A peptide and an absorbing agent for treating diseases related to anti-DNA carrier - useful for treatment of diseases related to anti-DNA antibodies and immune complexes.
XX Disclosure; Page 11, 14pp; Japanese.
XX The sequences given in AA57386 413 are peptides which are all covered

by the claimed generic formula:
 H-X (A-B)n-Y-Z
 A = Trp, Phe or a peptide fragment consisting of 2 residues;
 B = Trp, Phe, Asn or Glu;
 X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment consisting of 2-10 residues, provided that at least one of X or Y are present;
 Z = OH or NH₂; and
 n = 2-5.
 These peptides may be immobilized in a carrier in the preparation of an absorbing agent which may be used in the treatment of diseases related to anti-DNA antibodies and/or immune complex.
 Sequence 6 AA;
 Query Match 100.0%; Score 31; DB 15; Length 6;
 Best Local Similarity 50.0%; Pred. No. 7.90.05;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXXWKF 6
 DB 1 WVFWMF 6
 RESULT 2
 AAR57391 standard; Protein; 6 AA.
 AC AAR57391;
 XX 21-MAR 1995 (first entry)
 DE Trp, Phe or a peptide fragment consisting of 2 residues;
 KW Carrier, absorbing agent, treatment, anti-DNA antibody, immune complex.
 OS Synthetic.
 PN JF06122230-A
 PD 12-JUN-1994.
 PF 18-JAN 1993, 9307 00000000.
 PR 30-SEP-1992; 9207 0261821.
 PA (KWS) KURARAY CO LTD
 XX WPI, 1994 260510/32.
 DE A peptide and an absorbing agent, provided that at least one of a carrier, useful for treatment of diseases related to anti-DNA antibodies and immune complexes
 PT antibodies and immune complexes
 XX Discovered; Page 11; 14pp; Japanese.
 PS The sequence given in AAR57391 is the first entry which are all covered by the claimed generic formula:
 H-X-(A-B)n-Y-Z
 A = Trp, Phe or a peptide fragment consisting of 2 residues;
 B = Trp, Phe, Asn or Glu;
 X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment consisting of 2-10 residues, provided that at least one of X or Y are present;
 Z = OH or NH₂; and
 n = 2-5.
 These peptides may be immobilized on a carrier in the preparation of an absorbing agent which may be used in the treatment of diseases related to anti-DNA antibodies and/or immune complex.
 Sequence 6 AA;
 Query Match 100.0%; Score 31; DB 15; Length 6;

Best Local Similarity 50.0%; Pred. No. 7.90.05;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXXWKF 6
 DB 1 WVFWMF 6
 RESULT 3
 AAB01492
 ID AAB01492 standard; Peptide; 6 AA.
 AC AAB01492;
 XX 08-NOV-2000 (first entry)
 DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
 KW DNA binding, transcription factor, E2F, E2F 1, cell cycle, DP-1, activation, transactivation, apoptosis, proliferative disorder, psoriasis, restenosis.
 OS Synthetic.
 PN WO2000044771-A1.
 XX 03-AUG-2000.
 PF 26-JAN 2000; 2000WO 04002227.
 XX 26-JAN 1993, 9307 0001310.
 XX (FRO) FROLIFIX LTD.
 DE Mueller S, Hutterbach FE, Monclgland S;
 DR WPI; 2000-532806/48.
 XX Peptide which binds to the DNA binding domain of transcription factor E2F 1 and inhibiting cell cycle progression, useful for the treatment of cancer
 PT cancer
 XX Claim 6; Page 2; 4pp; English.
 CC Peptides which bind to the DNA binding domain of transcription factor E2F and inhibit cell cycle progression may be useful as research agents to investigate the interaction between E2F and DP-1, or the activation of transcription by E2F 1/DP 1 heterodimers. They may also be used for inducing apoptosis and/or cell cycle arrest in cells, particularly for treatment of cancer or other proliferative disorders such as psoriasis and restenosis.
 CC disorders such as psoriasis and restenosis.
 XX Sequence 6 AA;
 QY 1 WXXWKF 6
 DB 1 WVFWMF 6
 RESULT 4
 AAB01497
 ID AAB01497 standard; peptide; 6 AA.
 AC AAB01497;
 XX 08-NOV-2000 (first entry)
 DE Peptide which binds to transcription factor E2F 1 DNA binding domain.
 XX

KW DNA binding; transcription factor, E2F, E2F-1, cell cycle; DP-1,
KW activation; transcription, apoptosis, proliferative disorder,
XX psoriasis; restenosis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2 /note= "Any amino acid"
FT Misc-difference 3 /note= "Any amino acid"
FT Misc-difference 5 /note= "Any amino acid"
FT Misc-difference 5 /note= "Any amino acid"
XX
XX WO2000044771-A1.
XX
XX 03-AUG-2000
XX
XX 26-JAN-2000; 2000WO-GB00227.
XX
XX 26-JAN-1999; 99GB-0001710.
XX
XX (PROL-) PROLIFIX LTD
XX
XX Mueller R, Kontermann RE, Montigiani S;
XX WPI; 2000-532806/48
XX
XX Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
XX Claim 4; Page 9; 42pp; English
XX
XX Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 21; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXF 6
DB 1 WXXWXF 6

RESULT 5
AAB01499
ID AAB01499 standard; peptide; 6 AA.
XX
AC AAB01499;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW DNA binding; transcription factor; E2F, E2F-1, cell cycle; DP-1,
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Misc-difference 2 /note= "Any amino acid"
FT Misc-difference 3 /note= "Any amino acid"

FT /note= "Any amino acid"
XX
XX WO2000044771-A1.
XX
XX 03-AUG-2000
XX
XX 26-JAN-2000; 2000WO-GB00227
XX
XX 26-JAN-1999; 99GB-0001710
XX
XX (PROL-) PROLIFIX LTD.
XX
XX Mueller R, Kontermann RE, Montigiani S;
XX WPI; 2000-532806/48.
XX
XX Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of
PT cancer
XX
XX Claim 4; Page 9; 42pp; English.
XX
XX Peptides which bind to the DNA binding domain of transcription
CC factor E2F and inhibit cell cycle progression may be useful as
CC research agents to investigate the interaction between E2F and DP-1,
CC or the activation of transcription by E2F-1/DP-1 heterodimers. They
CC may also be used for inducing apoptosis and/or cell cycle arrest in
CC a cell, particularly for treatment of cancer or other proliferative
CC disorders such as psoriasis and restenosis.
XX
XX
SQ Sequence 6 AA;

Query Match 100.0%; Score 31; DB 21; Length 6;
Best Local Similarity 99.3%; Pred. No. 7.8e+05;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXF 6
DB 1 WXXWXF 6

RESULT 6
AAB01505
ID AAB01505 standard; peptide; 6 AA.
XX
AC AAB01505;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain
XX
KW DNA binding; transcription factor; E2F, E2F-1, cell cycle; DP-1;
KW activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
XX Synthetic.
OS
XX
XX WO2000044771-A1.
XX
XX 03-AUG-2000
XX
XX 26-JAN-2000; 2000WO-GB00227.
XX
XX 26-JAN-1999; 99GB-0001710.
XX
XX (PROL-) PROLIFIX LTD.
XX
XX Mueller R, Kontermann RE, Montigiani S;
XX WPI; 2000-532806/48.
XX
XX Peptides binding to the DNA binding domain of transcription factor E2F
PT and inhibiting cell cycle progression, useful for the treatment of

PT Cancer
XX
PS Example, Page 26, 42pp, English.
XX
CC Peptides which bind to the DNA binding domain of transcription factor E2F and inhibit cell cycle progression may be useful as research agents to investigate the interaction between E2F and DP-1, or the activation of transcription by E2F/DP-1 heterodimers. They may also be used for inducing apoptosis and/or cell cycle arrest in a cell, particularly for treatment of cancer or other proliferative disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;
Query Match 100.0%, Score 31, DB 21, Length 6;
Best Local Similarity 100.0%, Pred. No. 7.9e+05;
Matches 3, Conserved 3, Mismatches 0, Indels 0, Gaps 0;
OY 1 WXXWXF 6
DB 1 WARWHF 6
RESULT 7
AAB01506
ID AAB01506 standard; peptide; 6 AA.
XX
AC AAB01506;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding; transcription factor E2F, E2F-1, cell cycle; DP-1, activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN W0200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000W0-GP00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DH With, 2000 33256/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F and inhibiting cell cycle progression, useful for the treatment of cancer
XX
PS Example, Page 26, 42pp, English.
XX
CC Peptides which bind to the DNA binding domain of transcription factor E2F and inhibit cell cycle progression may be useful as research agents to investigate the interaction between E2F and DP-1, or the activation of transcription by E2F/DP-1 heterodimers. They may also be used for inducing apoptosis and/or cell cycle arrest in a cell, particularly for treatment of cancer or other proliferative disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;
Query Match 100.0%, Score 31, DB 21, Length 6;
Best Local Similarity 100.0%, Pred. No. 7.9e+05;
Matches 3, Conserved 3, Mismatches 0, Indels 0, Gaps 0;
OY 1 WXXWXF 6
DB 1 WARWHF 6

OY 1 WXXWXF 6
DB 1 WARWHF 6
RESULT 8
AAB01508
ID AAB01508 standard; peptide; 7 AA.
XX
AC AAB01508;
XX
DT 08-NOV-2000 (first entry)
XX
DE Peptide which binds to transcription factor E2F-1 DNA binding domain.
XX
KW DNA binding; transcription factor E2F, E2F-1, cell cycle; DP-1; activation; transcription; apoptosis; proliferative disorder;
KW psoriasis; restenosis.
XX
OS Synthetic.
XX
PN W0200044771-A1.
XX
PD 03-AUG-2000.
XX
PF 26-JAN-2000; 2000W0-GP00227.
XX
PR 26-JAN-1999; 99GB-0001710.
XX
PA (PROL-) PROLIFIX LTD.
XX
PI Mueller R, Kontermann RE, Montigiani S;
XX
DH With, 2000 33256/48.
XX
PT Peptides binding to the DNA binding domain of transcription factor E2F and inhibiting cell cycle progression, useful for the treatment of cancer
XX
PS Example, Page 26, 42pp, English.
XX
CC Peptides which bind to the DNA binding domain of transcription factor E2F and inhibit cell cycle progression may be useful as research agents to investigate the interaction between E2F and DP-1, or the activation of transcription by E2F/DP-1 heterodimers. They may also be used for inducing apoptosis and/or cell cycle arrest in a cell, particularly for treatment of cancer or other proliferative disorders such as psoriasis and restenosis.
XX
SQ Sequence 6 AA;
Query Match 100.0%, Score 31, DB 21, Length 6;
Best Local Similarity 100.0%, Pred. No. 7.9e+05;
Matches 3, Conserved 3, Mismatches 0, Indels 0, Gaps 0;
OY 1 WXXWXF 6
DB 1 WARWHF 6
RESULT 9
AAB12181
ID AAB12181 standard; peptide; 7 AA.
XX
AC AAB12181;
XX
DT 18-DEC-2001 (first entry)
XX
DE Polyglutamine repeat domain binding peptide, QP1-M7FD.
XX
KW Polyglutamine repeat domain; Huntington, atrophy; ataxia; androgen receptor; Friedreich's ataxia; dentatorubral pallidum atrophy; DRPLA;
KW Huntington's disease; dentatorubral pallidum atrophy; DRPLA;

KW spinobulbar muscular atrophy; spinocerebellar ataxia; neurotropic
 XX Unidentified.
 OS
 XX WO200168678-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 14-MAR-2001; 2001WO-0509222.
 PF
 XX 16-MAR-2000; 2000US 199791F.
 PR
 XX (UYDU-) UNIV DUKE.
 PA
 XX Strittmatter WJ, Burke JR, Nagai Y;
 PI
 XX WPI: 2001-414245/71
 DR
 XX
 PT New polypeptides which selectively bind to expanded polyglutamine
 PT repeat domains, useful for treating neurodegenerative diseases, e.g.
 PT Huntington's disease
 PS
 XX Example 13; Page 14; 59pp; English
 CC The present sequence is a polyglutamine repeat domain binding peptide.
 CC The peptide is useful for treating a cell that contains and expresses a
 CC protein having an expanded polyglutamine region (e.g. huntingtin;
 CC atrophin 1, ataxin 1, 2, 6 or 7, or androgen receptor protein), or
 CC treating a neurodegenerative disease characterised by the presence of
 CC expanded polyglutamine repeats, e.g. Huntington's disease, dentatorubral
 CC pallidum atrophy (DRPA), spinobulbar muscular atrophy, and
 CC spinocerebellar ataxia types 1, 2, 3, 6 and 7
 CC
 SQ Sequence 7 AA;
 OY 1 WXXWXP 6
 DB 1 WKWMPF 6
 X
 RESULT 10
 AAM45777
 ID AAM45777 standard; Peptide; 7 AA.
 AC
 XX AAM45777;
 AC
 XX 25-OCT-2001 (first entry)
 DT
 XX H11 binding site sequences containing peptide (CSP) #2049.
 DE
 XX Antigen-binding; tumour; diagnosis; stress protein-peptide complex; SPPC;
 KW immunogenically cross-reactive; cancer; immunogenic cancer cell;
 KW cytostatic; vaccine; tumour specific immunogenic response inducer;
 KW astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma, oligodendroglioma;
 KW ependymoma; medulloblastoma; primitive neural ectodermal tumour;
 KW
 OS Homo sapiens
 OS Synthetic.
 XX
 XX CA2290722-A1.
 PN
 XX 08-JUN-2001.
 PD
 XX 08-DEC-1999; 99CA-2290722
 PF
 XX 08-DEC-1999; 99CA-2290722.
 PR
 XX (NOVO) NOVOPHARM BIOTECH INC.
 PA
 XX

PI Kaplan HA, Maiti PK, East PJ, Herman W, Pan MD, Lewis KE;
 PI Entwistle JM, Macdonald GC;
 PP
 XX WPI: 2001 425937/46.
 DR
 XX
 PT Composition useful for treating and diagnosing cancer, comprising stress
 PT protein-peptide complexes associated with tumour, and isolated
 PT antigen-binding fragments of an antibody that binds specifically to the
 PT complex
 PS
 XX Example 4, Page 108, 154pp, English.
 CC The present invention describes a composition (1) comprising stress
 CC protein-peptide complexes (SPPC) associated with tumours that is
 CC specifically immunogenically cross-reactive with cell surface associated
 CC SPPCs specific to target cancer (TC). Also described is an isolated
 CC antigen binding fragment of an antibody that binds specifically to SPPCs
 CC or a population of different SPPCs consisting of immunogenic cancer cell
 CC surface associated SPPC of TC. (1) has cytotoxic activity and can be
 CC used in vaccine production and as a tumour-specific immunogenic response
 CC inducer. (1) is useful for treating 71 types of cancers or tumours in a
 CC subject, such as astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma,
 CC oligodendroglioma, ependymoma, medulloblastoma, and primitive neural
 CC ectodermal tumour (PNET). (1) is useful as cancer immunogen including
 CC vaccines. (1) is useful for diagnostic and palliative use, for detecting
 CC or imaging cancer cells, and to monitor the course of amelioration of
 CC malignancy in an individual. AAM45777 to AAM47109 represent peptides
 CC which are used in the exemplification of the present invention.
 CC
 SQ Sequence 7 AA;
 OY 1 WXXWXP 6
 DB 1 WPFWNP 6
 X
 RESULT 11
 AAR57399
 ID AAR57399 standard; Protein; 8 AA.
 AC
 XX AAR57399;
 AC
 XX 21-MAR-1995 (first entry)
 DT
 XX Peptide for treating diseases related to anti-DNA antibodies.
 DE
 XX Carrier; absorbing agent, treatment, anti-DNA antibody, immune response.
 KW
 OS Synthetic.
 OS
 XX JP26192290-A.
 PN
 XX 12-JUL-1994.
 PD
 XX 18-JAN-1993; 93JP-0006098.
 PF
 XX 30-SEP-1992; 92JP-0261921.
 PR
 XX (KIPS) KIPAPAY CO LTD.
 PA
 XX WPI: 1994 260510/32.
 DR
 XX A peptide and an adsorbing agent prep. by immobilising it on a
 PT carrier useful for treatment of diseases related to anti-DNA
 PT antibodies and immune complexes
 PS
 XX Disclosure; Page 12; 14pp; Japanese.
 CC The sequences given in AAR57386-413 are peptides which are all covered

by the claimed generic formula:
 H-X-(A-B)-Y-Z
 A = Trp, Phe or a peptide fragment consisting of 2 residues,
 B = Trp, Phe, Asn or Glu,
 X and Y = a bond or Asp, Glu, Arg, Tyr, His, a peptide fragment
 consisting of 2 to 10 residues, provided that at least one of
 X or Y are present;
 Z = OH or NH₂ and
 n = 2-5.
 These peptides may be formulated for use in the preparation of an
 absorbing agent which may be used in the treatment of diseases related
 to the immune system and/or immune defects.
 Sequence 8 AA;
 Query Match 100.0%; Score 31; DB 15; Length 8;
 Best Local Similarity 50.0%; Pred. No. 7.8e+05;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXXWKF 6
 DB 1 WXXWKF 6
 RESULT 12
 AAB01498 standard, Peptide, 9 AA.
 ID AAB01498 standard, Peptide, 9 AA.
 AC AAB01498,
 XX 08-NOV-2000 (first entry)
 DE Peptide which binds to transcription factor E2F1 DNA binding domain
 XX DNA binding, transcription factor, E2F, E2F1, cell cycle, CD 2,
 KW transcription, transcription, apoptosis, proliferation, cell cycle,
 KM proinvasive, metastasis.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 2 "Any amino acid"
 FT Misc-difference 3 "Any amino acid"
 FT Misc-difference 5 "Any amino acid"
 FT Misc-difference 7 "Any amino acid"
 FT Misc-difference 8 "Any amino acid"
 FT Misc-difference 8 "Any amino acid"
 XX
 PN W0200044771-A1.
 PD 03 AUG-2000.
 PF 26-JAN-2000; 2000WO 0800227.
 XX
 PR 26-JAN-1995, 35GB 000111.
 PA PROCTER & GAMBLE LTD.
 PI Mueller R, Kautzmann SE, Mottigand G
 DR WPI; 2000-032806/49.
 XX
 PT Peptides which bind to the DNA binding domain of transcription factor E2F1
 PT and inhibiting cell cycle progression, useful for the treatment of
 PI cancer
 XX
 PS Claim 4; Page 9; 42p; English
 XX
 CC Peptides which bind to the DNA binding domain of transcription

Factor E2F and inhibit cell cycle progression may be useful as
 CC research agents to investigate the interaction between E2F and Dp-1,
 CC a protein which is a transcription factor by E2F1/DP1 heterodimeric
 CC complex to regulate cell cycle progression and cell cycle arrest in
 CC a cell, particularly for treatment of cancer and for inhibiting
 CC disorders such as proinvasive and metastasis.
 XX
 SO Sequence 9 AA;
 Query Match 100.0%; Score 31; DB 15; Length 9;
 Best Local Similarity 50.0%; Pred. No. 7.8e+05;
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXXWKF 6
 DB 1 WXXWKF 6
 RESULT 13
 AAU93672 standard, Peptide, 9 AA.
 ID AAU93672 standard, Peptide, 9 AA.
 AC AAU93672;
 XX 02-JUN-2002 (first entry)
 DE Granulocyte colony stimulating factor receptor binding peptide #498.
 XX
 KW 3-OSF, granulocyte colony stimulating factor receptor, cytokines;
 KW haematopoietic growth factor; neutrophil proliferation; AIDS;
 KW neutrophil differentiation; acquired immunodeficiency syndrome;
 KW chemotherapy induced neutropenia; chemotherapy acquired pneumonia;
 KW depressed neutrophil count; immunostimulant.
 XX
 CC Synthetic.
 XX
 RN W0200207676 A2.
 XX
 PD 31 JAN-2002.
 PF 20 JAN 2002; 2001WO 0572046.
 XX
 PR 20 JAN 2002, 200000 000031.
 XX
 PA 35AA; 35AA; 35AA LTD.
 PI Gwira SE, Balu P, Duffin DJ, Pipani S, McEwen-Merrill B;
 PI Schatz PJ;
 XX
 BK WPI; 2002 000000/00.
 XX
 PT Novel compounds, useful for treating depressed neutrophil count,
 PT comprising peptide chains of approximately 6 to 40 amino acids in length
 PT that bind to granulocyte colony stimulating factor receptor -
 XX
 PS Claim 4; Page 67; 90pp; English.
 XX
 CC The invention relates to compounds comprising a peptide chain
 CC approximately 6 to 40 amino acids in length that binds to granulocyte-
 CC colony stimulating factor receptor (G-CSF). The compounds contain
 CC specific sequences of the generic peptides appearing as AAU93672-AAU93676
 CC and the generic sequences XY1XY2XY3XY4XY5XY6XY7XY8XY9 (where
 CC XY1 = E, G, Q, V or Y; XY2 = E, A, L, K, E, W or G; XY3 = Y, R or G;
 CC XY4 = L, A or V; XY5 = R, A, M, H, E, V, L, G, D, Q or S; XY6 = E or
 CC Y; XY7 = A or G; and XY8 = R, H, G or L; and XY9 = A, S or D;
 CC EXVI EXVI EXVI EXVI 9 (where EXVI = A, E or G; EXVI2 = F, H or D;
 CC EXVI3 = F, G, I, V, T, L, F, V, W, R, K, L, Q, A or G; EXVI4 = R
 CC or E; and EXVI5 = W, G or R). The compounds are used for treating
 CC conditions associated with depressed neutrophil count, e.g. chemotherapy
 CC induced neutropenia, and induced neutropenia in generally acquired
 CC pneumonia-induced pneumonia. The compounds are useful as in vitro as
 CC tools for understanding the biological role of granulocyte colony

CC stimulating factor (G-CSF a haematopoietic growth factor and
CC cytokine that stimulates neutrophil proliferation and differentiation),
CC including evaluation of many factors thought to influence, and be
CC influenced by, production of white blood cells, in the development of
CC compounds that bind to G-CSFR, as reagents for detecting G-CSF receptor
CC or related receptor on living cells, fixed cells, in biological fluid, in
CC tissue homogenates or in purified natural biological materials, in situ
CC staining, fluorescence activated cell sorting (FACS), Western blotting or
CC enzyme-linked immunosorbent assay (ELISA), in receptor purification or
CC in purifying cells expressing G-CSFR on the cell surface (or inside
CC permeabilised cells) as a commercial research reagent for various medical
CC and diagnostic uses or to treat a disease that would benefit from the
CC ability to of a compound to mimic the effects of G-CSF in vivo.
CC The compounds bind specifically to G-CSFR and allow for studies of
CC biological activities mediated by the receptor and for the treatment of
CC diseases, disorders and conditions that would benefit from activating or
CC inactivating G-CSFR. The present sequence is a G-CSFR binding peptide of
CC the invention.

XX Sequence 9 AA;

Query Match 100.0%; Score 31; DB 23; Length 9;
Best Local Similarity 100.0%, Pred. No. 7 Acc. 97,
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|:
Db 4 WWMWNF 9

RESULT 14

AA057400
ID AAR57400 standard; Protein; 10 AA

XX AAR57400;

DT 21-MAR-1995 (first entry)

DE Peptide for treating diseases related to anti-DNA antibodies

KW Carrier, absorbing agent, treatment, anti-DNA antibody, 3p, 1000-100000

XX Synthetic

PN JP06192290-A.

XX 12 JUL 1994.

PF 18-JAN-1993; 93JP-0006098.

PR 30-SEP-1992; 92JP 0261921.

PA (KUPS) KUPAPAY CO LTD.

XX WPI; 1994-260510/32

XX A peptide and an adsorbing agent prepd by immobilising it on a

PT carrier - useful for treatment of diseases related to anti-DNA

PS antibodies and immune complexes

XX Disclosure; Page 12; 14pp; Japanese

CC The sequences given in AAR57400 413 are peptides which are all covered

CC by the claimed generic formula:
H-X-(A-B)n-Y-Z

CC A = Trp, Phe or a peptide fragment consisting of 2 residues;
CC B = Trp, Phe, Asn or Glu;

CC X and Y = a bond or Asp, Glu, Arg, Lys, His or a peptide fragment
CC consisting of 2-10 residues, provided that at least one of
CC X or Y are present;

CC Z = OH or NH2, and
CC n = 2-5.
CC These peptides may be immobilised on a carrier in the preparation of an

CC absorbing agent which may be used in the treatment of diseases related
CC to anti-DNA antibodies and/or immune complex.

XX Sequence 10 AA;

Query Match 100.0%; Score 31; DB 15; Length 10;
Best Local Similarity 100.0%, Pred. No. 1.3e+02,
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXXWXP 6
|:|:|:
Db 3 WWMWNF 8

RESULT 15

AA062149
ID AAW62149 standard; peptide, 11 AA.

XX AAW62149;

DT 16-SEP-1998 (first entry)

DE Haemophilus influenzae tyrosine tRNA synthetase binding peptide 22.

KW Identification; ligand; biological activity; target-binding;
XX drug screening; library; inhibitory ligand.

XX Synthetic.

OS Haemophilus influenzae.

PN WO9819162-A1.

XX 07-MAY-1998.

PF 31-OCT-1997; 97WO-US19638.

PR 31-OCT-1996; 96US-0740671.

PA (NOVA-) NOVALON PHARM CORP.

XX Fowlkes TM, Frolinger JA, Hydro-Tetrasider PE, Kay PK;

XX WPI; 1998-272389/24.

XX Identifying ligands which mediate biological activity of a protein
PT by identifying target-binding ligands and screening a library for
PT ligands which inhibit target-binding ligand mediated activity

PS Example 5; Page 100; 143pp; English.

XX A method has been developed for identifying a ligand which mediates the
CC biological activity of a target protein (T) by inhibiting the binding
CC of (T) to a binding partner. The method comprises: (a) screening a first
CC combinatorial library comprising first member ligands for binding to the
CC target binding ligands (TBLs), to identifying one or more TBLs; (b)
CC screening a second library comprising second member ligands for the
CC ability to inhibit the binding of one or more of the TBLs to the target
CC protein, and so obtaining one or more inhibitory ligands; and (c)
CC determining which of the inhibitory ligands can mediate a biological
CC activity of the target protein. The present sequence represents a
CC potential binding peptide for Haemophilus influenzae tyrosine tRNA
CC synthetase from an example of the present invention. The method can be
CC used for identifying drugs which can mediate the biological activity of
CC a target protein. It can be used to identify the biological activity of
CC a target protein whose biological function is not known and perhaps
CC cannot be determined directly. The method can also be used to identify
CC new inhibitory ligands of specific target proteins. The method provides
CC high throughput screens which are essentially identical for similar and
CC dissimilar targets, bypassing the need to develop distinct assays for
CC biochemically diverse targets.

XX Sequence 11 AA;

Query Match 100.0%; Score 31; DB 19; length 11;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

CY 1 WXXWXP 6
DB 5 WPDWGP 10

Search completed: February 20, 2003, 10:47:36
Job time : 31.333 secs